

Univerzita Karlova v Praze
Filozofická fakulta
Ústav informačních studií a knihovnictví

Studijní program: Informační studia a knihovnictví

Studijní obor: Informační studia a knihovnictví

Kristýna Kožuchová

**Polytematický strukturovaný heslář (PSH) a Předmětová hesla Kongresové
knihovny (LCSH) v kontextu vybraných trendů ve vývoji systémů organizace
znalostí**

**Polythematic Structured Subject Heading System (PSH) and Library of Congress
Subject Headings (LCSH) in the context of trends in Knowledge Organization
Systems**

Diplomová práce

Praha 2011

Poděkování

Za pomoc při přípravě, odborné vedení a dohled, cenné rady, připomínky a vstřícnost při vedení této diplomové práce děkuji především Mgr. Kláře Havlíčkové. Dále bych ráda poděkovala Mgr. Marii Balíkové, která byla vstřícná a pomohla mi při verifikaci překladu termínu Knowledge Organization Systems. V neposlední řadě bych ráda poděkovala celému referátu PSH a jeho vedení v Národní technické knihovně, bez nichž by tato práce patrně nevznikla.

Vedoucí diplomové práce:

Mgr. Klára Havlíčková

Oponent diplomové práce:

Datum obhajoby:

Hodnocení:

Prohlášení:

Prohlašuji, že jsem diplomovou práci vypracovala samostatně, že jsem řádně citovala všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze, dne 18. srpna 2011

.....

Identifikační záznam

KOŽUCHOVÁ, Kristýna. *Polytematický strukturovaný heslář (PSH) a Předmětová hesla Kongresové knihovny (LCSH) v kontextu vybraných trendů ve vývoji systémů organizace znalostí [Polythematic Structured Subject Heading System (PSH) and Library of Congress Subject Headings (LCSH) in the context of trends in Knowledge Organization Systems]*. Praha, 2010. 82 s. Diplomová práce. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví. Vedoucí diplomové práce Mgr. Klára Havlíčková.

Abstrakt

Diplomová práce analyzuje Polytematický strukturovaný heslář (PSH) a Předmětová hesla Kongresové knihovny (LCSH) v kontextu vybraných trendů ve vývoji systémů organizace znalostí. Za tímto účelem jsou nejprve představeny systémy organizace znalostí (KOS), jejich taxonomie a návrh způsobu jejich hodnocení. Dále je charakterizován World Wide Web v kontextu organizace znalostí se zaměřením na webovou éru KOS. Třetí kapitola se věnuje vybraným trendům ve vývoji KOS, konkrétně se věnuje interoperabilitě KOS a aplikaci webových technologií. Ve čtvrté kapitole jsou na základě navrženého způsobu hodnocení KOS popsány dva tradiční zástupci KOS, tedy PSH a LCSH. Poslední kapitola se zaměřuje na analýzu vybraných trendů ve vývoji KOS v kontextu PSH a LCSH. Analýza představuje detailnější pohled na převod PSH a LCSH do datového modelu SKOS, který je společný oběma systémům, stejně jako uvádí příklady interoperability těchto systémů.

Klíčová slova

Polytematický strukturovaný heslář, Předmětová hesla Kongresové knihovny, PSH, LCSH, systémy organizace znalostí, KOS, organizace znalostí, SKOS, interoperabilita KOS, WWW, World Wide Web, Simple Knowledge Organization System

OBSAH

OBSAH	6
PŘEDMLUVA	8
ÚVOD	10
1. SYSTÉMY ORGANIZACE ZNALOSTÍ.....	12
1.1. TAXONOMIE SYSTÉMŮ ORGANIZACE ZNALOSTÍ.....	13
1.1.1. <i>Rejstříky termínů</i>	18
1.1.2. <i>Klasifikace a kategorie</i>	19
1.1.3. <i>Rejstříky vztahů</i>	20
1.2. ZPŮSOBY HODNOCENÍ SYSTÉMŮ ORGANIZACE ZNALOSTÍ	20
1.2.1. <i>Vnitřní parametry hodnocení KOS</i>	21
1.2.2. <i>Vnější parametry hodnocení KOS</i>	22
1.3. PODPORA VÝZKUMU A VÝVOJE V OBLASTI SYSTÉMŮ ORGANIZACE ZNALOSTÍ.....	22
1.3.1. <i>Mezinárodní organizace ISKO</i>	22
1.3.2. <i>Síť NKOS</i>	23
2. VLIV WEBOVÉHO PROSTŘEDÍ NA SYSTÉMY ORGANIZACE ZNALOSTÍ.....	25
2.1. CHARAKTERISTIKA WEBU V KONTEXTU ORGANIZACE INFORMACÍ	25
2.2. PŘED WEBOVÁ ÉRA KOS	30
2.3. WEBOVÁ ÉRA KOS.....	31
3. VYBRANÉ TRENDY SYSTÉMŮ ORGANIZACE ZNALOSTÍ	34
3.1. INTEROPERABILITA.....	36
3.1.1. <i>ISO: 25964</i>	38
3.2. APLIKACE WEBOVÝCH TECHNOLOGIÍ	40
3.2.1. <i>SKOS</i>	40
4. TRADIČNÍ SYSTÉMY ORGANIZACE ZNALOSTÍ NA PŘÍKLADU PSH A LCSH	43
4.1. PSH.....	43
4.1.1. <i>Charakteristika PSH podle navrženého modelu – vnitřní vlastnosti</i>	44
4.1.1.1. <i>STRUKTURA, ENTITY A VZTAHY PSH</i>	44
4.1.1.2. <i>STANDARDIZACE PSH</i>	45
4.1.1.3. <i>JAZYK PSH</i>	45
4.1.1.4. <i>DOMÉNA PSH</i>	47
4.1.1.5. <i>MÉDIA A ZOBRAZENÍ PSH</i>	47
4.1.2. <i>Charakteristika PSH podle navrženého modelu – vnější vlastnosti</i>	49

4.1.2.1.	ÚČEL PSH.....	49
4.1.2.2.	KOMUNITA UŽIVATELŮ PSH.....	50
4.1.2.3.	AKTUALIZACE PSH	50
4.2.	LCSH.....	53
4.2.1.	Charakteristika LCSH podle navrženého modelu – vnitřní vlastnosti	54
4.2.1.1.	STRUKTURA, ENTITY A VZTAHY LCSH.....	54
4.2.1.2.	STANDARDIZACE LCSH.....	55
4.2.1.3.	JAZYK LCSH	55
4.2.1.4.	DOMÉNA LCSH	57
4.2.1.5.	MÉDIA A ZOBRAZENÍ LCSH	57
4.2.2.	Charakteristika LCSH podle navrženého modelu – vnější vlastnosti.....	60
4.2.2.1.	ÚČEL LCSH	60
4.2.2.2.	KOMUNITA UŽIVATELŮ LCSH.....	61
4.2.2.3.	AKTUALIZACE LCSH	61
5.	ANALÝZA VYBRANÝCH TRENDŮ VE VÝVOJI SYSTÉMŮ ORGANIZACE ZNALOSTÍ NA PŘÍKLADECH PSH A LCSH	64
5.1.	PSH A VYBRANÉ TRENDY VE VÝVOJI KOS	64
5.1.1.	PSH v datovém modelu SKOS.....	64
5.1.2.	Interoperabilita PSH.....	67
5.2.	LCSH A VYBRANÉ TRENDY VE VÝVOJI KOS	69
5.2.1.	LCSH v datovém modelu SKOS	69
5.2.2.	Interoperabilita LCSH	71
5.3.	LOD CLOUD DIAGRAM	72
	ZÁVĚR	73
	SEZNAM POUŽITÉ LITERATURY	75
	SEZNAM POUŽITÝCH ZKRATEK	81
	SEZNAM OBRÁZKŮ	82

PŘEDMLUVA

Tématem předkládané diplomové práce je analýza vybraných trendů ve vývoji systémů organizace znalostí na příkladech Polytematického strukturovaného hesláře a Předmětových hesel Kongresové knihovny.

Jedním z cílů této práce je snaha porozumět komplexnosti problematiky taxonomie, způsobu hodnocení systémů organizace znalostí, především uvědomění si potenciálu, který v sobě tyto systémy skrývají, a který může být využit nejen v knihovnách, ale s využitím vhodných nástrojů také v prostředí World Wide Web.

K podnětům, které vedly ke vzniku této práce patří jednak pracovní zkušenost autorky v referátu PSH v Národní technické knihovně, kde se od roku 2008 věnuje správě, údržbě a propagaci Polytematického strukturovaného hesláře, jednak její zájem o systémy organizace znalostí a jejich uplatnění v prostředí World Wide Web.

S terminologickým souslovím *systémy organizace znalostí* (z angličtiny *Knowledge Organization Systems*) se v českém prostředí, až na výjimky, téměř nesetkáme, neboť zde není zavedeno, a tudíž není ustáleno jako český ekvivalent. Ani v České terminologické databázi knihovnictví a informační vědy (TDKIV) tento termín či jiné ekvivalenty nenalezneme, přestože se již několik let objevuje v zahraniční literatuře. Sporadicky se v některých pracích jednotlivců sice objevuje, avšak v jiných překladech, než jaký byl pro tuto práci zvolen, na příklad jako sousloví *systémy pořádání informací a znalostí* nebo *systémy pro pořádání poznatků*. Tyto překlady pravděpodobně vychází z ustálenějšího překladu pro anglický termín *knowledge organization* – pořádání znalostí. Termín pořádání znalostí byl více či méně přijat jako český ekvivalent k tomu anglickému, zatímco jeho doslovnější překlad *organizace znalostí* zůstal více v pozadí. Proč jsem se tedy nepřiklonila k již zavedenému, v českém knihovnictví používanému termínu pořádání znalostí? Návrh překladu byl po dohodě s vedoucí práce konzultován s Mgr. Marií Balíkovou, která je nejen vedoucí Oddělení národních věcných autorit a věcného zpracování v Národní knihovně České republiky, ale která má za sebou i bohatou publikační činnost z oblasti věcného zpracování dokumentů. Po telefonické dohodě, která proběhla na podzim roku 2010, jsme se dohodly, že v této diplomové práci mohu použít překladu *systémy organizace znalostí* raději než *systémy pořádání*

znalostí. Důvodem je vývoj v terminologické oblasti, a tudíž jeho stále častější a modernější podoba.

Tato diplomová práce je rozčleněna do pěti kapitol a dalších podkapitol, jejichž součástí mohou být obrázky. První kapitola je věnována systémům organizace znalostí, jejich taxonomii a způsobu hodnocení. Druhá kapitola zachycuje některé podstatné charakteristiky World Wide Web ve vztahu k organizaci znalostí, stejně jako stručně charakterizuje webovou éru systémů organizace znalostí. S vlivem webu úzce souvisejí také trendy, které lze vysledovat ve vývoji systémů organizace znalostí. Proto je následující třetí kapitola věnována právě vybraným trendům. Čtvrtá a pátá kapitola zahrnuje vlastní rozbor a analýzu PSH a LCSH jako systémů organizace znalostí v kontextu vybraných trendů popsaných ve třetí kapitole.

Pro citování dokumentů je v této práci využit tzv. harvardský systém citování v souladu s normami ČSN ISO 690 a ČSN ISO 690-2. Dokumenty jsou v seznamu použité literatury řazeny abecedně podle záhlaví.

ÚVOD

Klasifikační schémata, slovníky, tezaury, předmětová hesla, ale i různé autoritní soubory a jiné typy systémů organizace znalostí jsou v knihovnách vytvářeny už celá desetiletí. Jsou to tedy právě knihovny, které za sebou mají dlouholetou historii týkající se tvorby, implementace, užití a poskytování systémů organizace znalostí, které jsou klíčovým mechanismem pro organizaci znalostí jak v těchto knihovnách, ale také v jiných paměťových a informačních institucích, tak v dalších relevantních subjektech či systémech.

Prostředí, ve kterém tyto systémy vznikly se však velmi změnilo, stejně jako se proměnila společnost. V posledních deseti letech došlo k neobyčejnému technologickému pokroku, objevily se nové fenomény doby, příkladem za všechny je sociální síť Facebook nebo Twitter. Zároveň došlo k velkému rozvoji Internetu a začalo se hovořit o konceptu webu 2.0. Podle mého názoru, všechny tyto faktory a mnohé další, proměnily nejen postavení knihoven a prostředí, ve kterém se nacházejí, ale změnily i pohled na vývoj a aplikaci systémů organizace znalostí.

Dříve byly systémy organizace znalostí v podstatě izolovány ve svém mateřském prostředí, zatímco s nástupem počítačů, vznikem Internetu a World Wide Web a dalším rozvojem informačních a počítačových technologií se tvůrcům těchto systémů otevřely nové možnosti pro jejich využití. Systémy organizace znalostí mohou díky tomuto rozvoji, ale také díky snaze komunity sémantického webu a různých metadatových iniciativ, rozvinout svůj potenciál a prokázat svou sofistikovanost. Otázkou tedy je, jakým způsobem mohou být tyto nástroje, které byly a jsou efektivní v bibliografickém informačním prostředí využity v prostředí World Wide Web.

Polytematický strukturovaný heslář (PSH) vytvářený v Národní technické knihovně a Předmětová hesla Kongresové knihovny (LCSH) jsou představiteli tradičních systémů organizace znalostí a jako takové slouží primárně k předmětové indexaci dokumentů, ať už ve svých mateřských knihovnách nebo v dalších spolupracujících knihovnách a institucích.

Ačkoli se může zdát, že jsou tyto systémy z určitého pohledu neporovnatelné, odlišné, že nemohou mít nic společného, není tomu tak. Přestože PSH a LCSH mají za sebou oba různě dlouhou historii. LCSH jsou oproti PSH nesrovnatelně starší. Vždyť vývoj

LCSH probíhá již od počátku minulého (20. století), zatímco PSH je na světě teprve necelých dvacet let. Cestu, kterou LCSH od té doby urazil, byla nesmírně důležitá nejen pro samotnou Kongresovou knihovnu, ve které a pro kterou LCSH původně vznikl, ale jeho vývoj byl směrodatný pro mnoho dalších knihoven, které jej postupně začaly využívat. Jeho vliv je patrný také u vývoje zásad pro tvorbu a aplikaci předmětových hesel.

Jakkoli se proměnilo prostředí, ve kterém tyto systémy vznikly, je u jejich tvůrců zřetelná snaha o to, aby byly jak PSH, tak LCSH, dále využitelné, a to nejen v tradičním prostředí knihoven a jiných institucích, ale i v rámci World Wide Web. Tato snaha, společná oběma systémům, více či méně předurčila vznik této diplomové práce.

I Přes rozdílné stáří, tradici, komunitu uživatelů, autoritu a personální sílu obou systémů věřím, že za tu poměrně krátkou dobu se PSH v určitém ohledu dokázal LCSH, a to i přes jejich četné rozdíly, vyrovnat.

1. SYSTÉMY ORGANIZACE ZNALOSTÍ

Systémy organizace znalostí jsou systémy, které, zjednodušeně řečeno, nějakým způsobem organizují znalosti. Organizace znalostí (Knowledge Organization, KO)¹ má podle Hjørlanda [2008] dva významy, užší a širší. V užším slova smyslu je organizace znalostí o činnostech jako je popis dokumentu, indexace a klasifikace vykonávaných v knihovnách, bibliografických databázích, archivech a dalších paměťových institucích. Tyto činnosti provádějí knihovníci, archiváři, předmětoví specialisté, ale i počítačové algoritmy nebo laici. Organizace znalostí tak může být chápána nebo definována jako *„Organizace a reprezentace zaznamenaných znalostí s cílem zajistit efektivní využívání údajů, informací a poznatků obsažených v informačních zdrojích. Pořádání informací zahrnuje vytváření organizovaných souborů identifikačních a obsahových informací získaných informační analýzou dokumentu podle předem stanoveného systému. Podle typu vstupních informací se vyčleňuje identifikační pořádání informací a věcné pořádání informací“* [BALÍKOVÁ, 2003]. V širším významu je organizace znalostí o sociálním rozdělení duševní práce. Jedná se na příklad o organizaci univerzit a dalších výzkumných institucí zaměřených na vyšší vzdělávání, strukturu disciplín a profesí nebo produkci a šíření znalostí². Širší význam je jednak o tom, jak jsou znalosti organizovány sociálně, jednak o tom, jak je organizována samotná realita.

Zatímco knihovní a informační věda se zabývá organizací znalostí v užším slova smyslu, ostatní disciplíny, jako na příklad sociologie znalostí, studují organizaci znalostí v rámci širšího pojetí tohoto termínu. Hjørland [2008] vyjádřil názor, že knihovní a informační věda často ignorovala širší význam organizace znalostí. Správné porozumění problematice organizace znalostí se však bez pochopení širšího pojetí neobejde.

Organizace znalostí jako předmět studia se zabývá podstatou a kvalitou procesu organizace znalostí (Knowledge Organizing Processes, KOP), stejně jako se zabývá systémy organizace znalostí, které se užívají k organizaci dokumentů.

¹ V českém prostředí se častěji setkáme s terminologickým ekvivalentem *pořádání znalostí* nebo *pořádání informací*.

² Příkladem studie, která se zabývá organizací znalostí v širším významu je podle Hjørlanda [2008] dílo A. Olesona *The Organization of knowledge in modern America, 1860-1920* vydaného v roce 1979.

Termín *systémy organizace znalostí* (Knowledge Organization Systems³, KOS) byl vytvořen pracovní skupinou *Networked Knowledge Organization Systems Working Group* v rámci setkání na konferenci ACM Digital Libraries konané ve dnech 24. – 27. června 1998 v Pittsburghu (Pensylvánie) [HODGE, 2000]. Tradičně jsou systémy organizace znalostí (dále jen KOS) vytvářeny a užívány v knihovnách, kde slouží k organizaci materiálů za účelem jejich vyhledávání. Podle Hodgeové [2000] si je lze představit jako mosty mezi informační potřebou uživatele a dokumenty či materiály v kolekci. Proto by uživatel měl být schopen identifikovat předmět svého zájmu bez předchozího vědomí o jeho existenci. „*Termín systémy organizace znalostí postihuje všechny typy schémat, které slouží k organizaci informací a podporující znalostní management. Systémy organizace znalostí zahrnují klasifikační schémata, která organizují materiály na obecné úrovni, předmětová hesla, která poskytují detailnější přístup a soubory autorit, jejichž cílem je kontrola různých variant verzí klíčových informací (např. geografická či osobní jména). K těmto systémům se ale také řadí méně tradiční schémata jako jsou sémantické sítě a ontologie*“ [HODGE, 2000].

KOS jsou mechanismem pro organizaci informací, a proto jsou srdcem každé knihovny, muzea a archivu [HODGE, 2000]. K tradičním zástupcům KOS patří na příklad Třídění Kongresové knihovny, Deweyho desetinné třídění nebo Mezinárodní desetinné třídění, jež spojuje potřeba uchovat jednotlivé jednotky na určitém místě na regále. K poskytnutí více kritériálního přístupu, který zachází za limity pouhé fyzické lokace jednotky, se používají předmětová hesla. Na úrovni specifického obsahu pak vytvářejí knihovny soubory autorit pro kontrolu osobních, korporativních a geografických jmen.

Znalost KOS a jejich charakteristik je nezbytná pro návrh jakéhokoli systému pro vyhledávání informací nebo systémů báze znalostí, především je pak nepostradatelná v digitálním prostředí [SOUZA; TUDHOPE; ALMEIDA, 2011].

1.1. Taxonomie systémů organizace znalostí

Nezáleží na tom, jak rozsáhlé jsou jednotlivé taxonomie KOS, každá snaha vyjmenovat všechny systémy dříve nebo později narazí na zrak jiného kritického pozorovatele. To bude trvat tak dlouho, dokud různé interpretace toho, co nazýváme KOS, povede k různým výsledkům [SOUZA; TUDHOPE; ALMEIDA, 2011]. Souza; Tudhope;

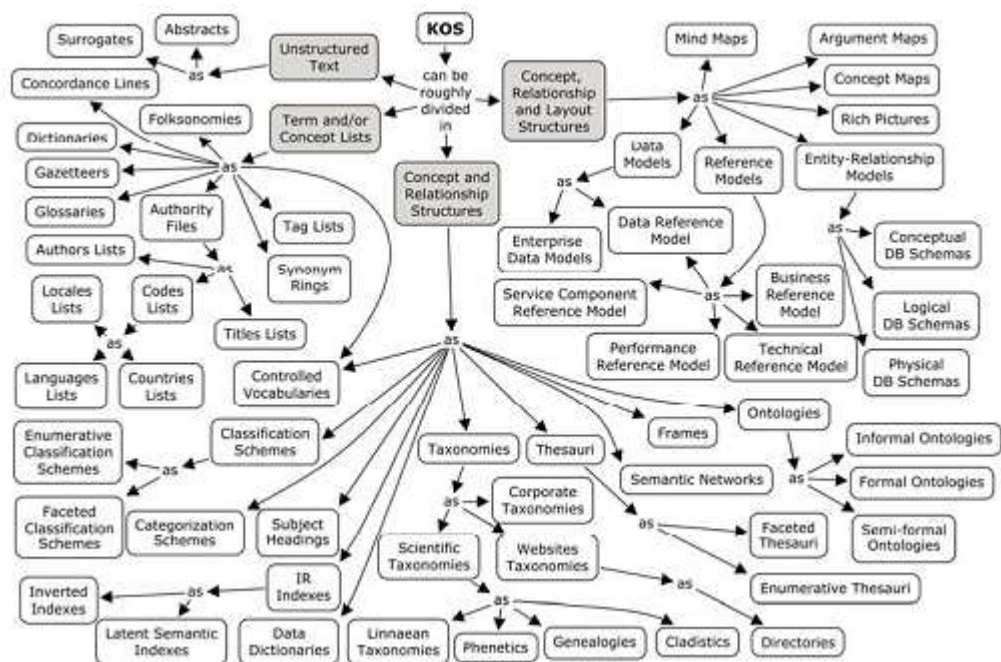
³ Hjørland [2008] užívá spojení *knowledge organizing systems*.

Almeida [2010] konstatují: „...jsme vzdáleni konsenzu o taxonomii systémů organizace znalostí a příbuzné terminologii.“ Tak, jak se ukazuje v mnoha jiných vědních oborech a disciplínách, i na poli knihovní a informační vědy se setkáváme s problémem terminologie. Různé komunity či skupiny definující KOS užívají různou terminologii pro stejné koncepty a naopak stejné termíny pro různé koncepty. Problematická terminologie a nejednotnost tak vede k neurčitosti skryté ve formě polysémie a synonymie.

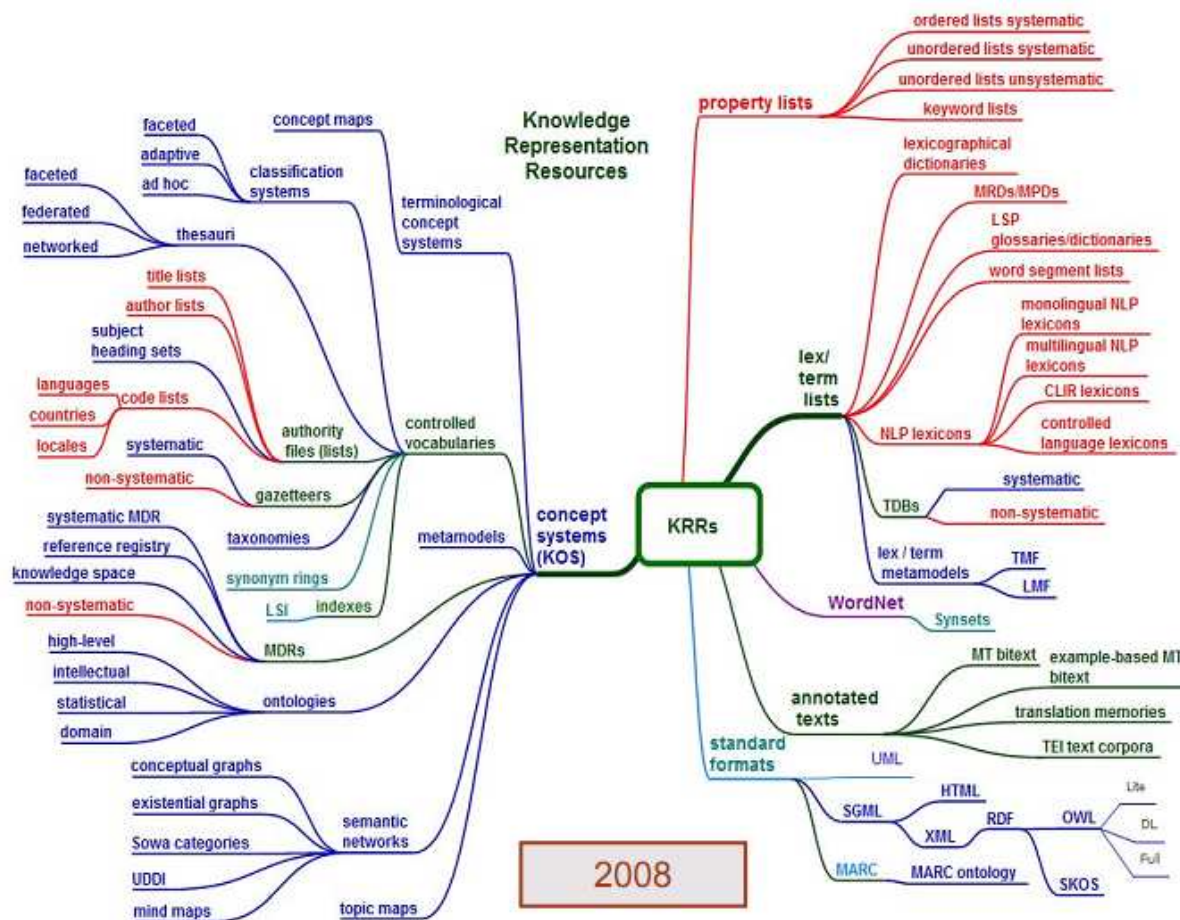
Existuje mnoho hledisek, podle kterých mohou být KOS hodnoceny a na základě nich rozděleny. Mnoho z těchto aspektů nelze vnímat izolovaně, jelikož jsou navzájem závislé. Celá řada autorů, kteří se touto problematikou zabývali nebo zabývají, zkoumá KOS z různých pohledů. Mezi ně patří na příklad Gaile Hodge [2000], Sue Ellen Wright [2008], Marcia Lei Zeng [2008], Renato Rocha Souza, Doug Tudhope a Maurício Barcellos Almeida [2010]⁴.

Zatímco v návrhu taxonomie KOS Souzy, Tudhopea a Almeidy [2010] (viz obr. č. 1) je hlavním kritériem rozdělení jednotlivých KOS typ struktury rozkládající se od nestrukturovaných textů (Unstructured Text) až k těm, které pohlíží na koncepty, vztahy a prostorové rozložení (Concept, Relationship and Layout Structures) jako na součást struktury, Wright [2008] rozlišuje stupně systematickosti KOS (viz obr. č. 2). Systematické zdroje jsou zobrazeny modře (od mělkých struktur až po ty hluboké, sémantické). Červená barva znázorňuje nesystematické zdroje, především rejstříky s náhodnými nebo tradičními principy organizace. Zelená barva představuje hybridní nadřazené uzly, které zahrnují obojí, systematické a nesystematické zdroje. Wordnet s vnitřním hybridním systémem, plytkou systematikou a lexikografickým přístupem je znázorněn fialově.

⁴ Nejedná se o vyčerpávající výčet osobností, které se problematice taxonomie KOS věnují. K dalším autorům patří na příklad Dagobert Soergel.



Obr. 1: Taxonomie KOS podle Souzy; Tudhope; Almeida [2010]



Obr. 2: Taxonomie KOS a zdroje reprezentace znalostí podle Wright [2008]

Ačkoli je taxonomie KOS Souzy, Tudhopea a Almeidy [2010] podrobná, je podobně jako taxonomie KOS podle Wright [2008] poměrně složitá. Nevýhodou této taxonomie KOS je také skutečnost, že sami autoři přiznávají, že se jedná o návrh, který bude pravděpodobně dále rozpracován, tudíž lze očekávat v budoucnu změny a doplňky. Tato taxonomie navíc na rozdíl od vybrané taxonomie KOS podle Hodgeové [2000] postrádá jakýkoli konkrétní příklad k jednotlivým skupinám, což je prvek, který znesnadňuje celkové porozumění navržené taxonomie.

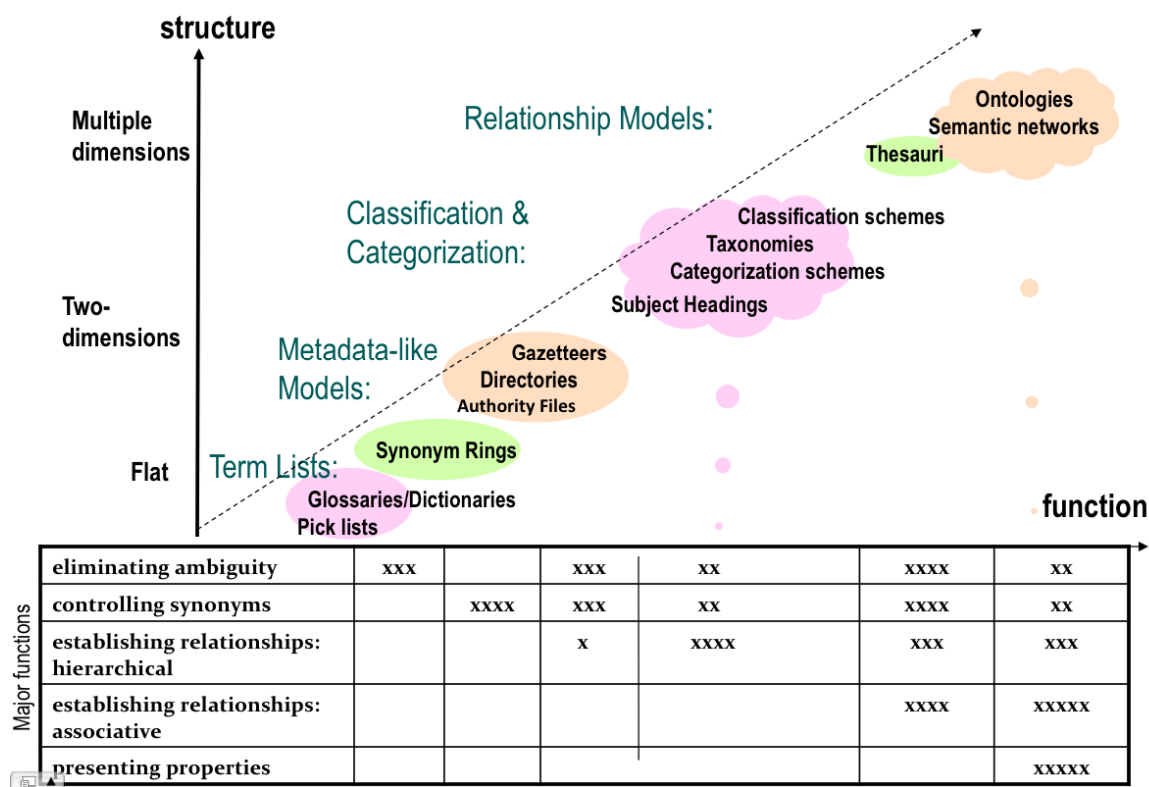
Z taxonomie KOS podle Wright [2008] je patrné, že autorka v rámci jednoho schématu znázorňuje nejen různé typy KOS, ale i zdroje reprezentace znalostí (Knowledge Representation Resources, KRRs), které střídavě označuje také jako schémata organizace znalostí (Knowledge Organization Schemes), zároveň zachycuje formáty, které slouží k jejich reprezentaci (např. MARC, SGML, HTML apod.). Schéma proto může působit neuspořádaně a nesourodě.

Pro podrobnější rozbor byla vybrána taxonomie KOS podle Hodgeové [2000]. Tato taxonomie byla zvolena z několika důvodů. Jedním z nich je fakt, že vybraná taxonomie poskytuje základní přehled a rozdělení KOS spolu s konkrétními příklady doplněnými jejich stručným popisem. Díky tomu je taxonomie přehledná a srozumitelná. Navíc byla přijata skupinou NKOS, významnou komunitou odborníků v oblasti KOS (viz níže).

Vizuální sumarizace taxonomie KOS založené na rozdělení KOS podle Hodgeové [2000], která byla rovněž přijata skupinou NKOS [NETWORKED, 2000], je zobrazena na obr. č. 3. Tato taxonomie, jejíž autorem je Marcia Lei Zeng, zachycuje různé typy KOS, které jsou sestaveny podle stupně kontroly (od přirozeného jazyka až k jazyku umělému) a síly jejich sémantické struktury (od slabých struktur k těm silným) s ohledem na obecné hlavní funkce těchto systémů. Z tohoto pohledu se k nejméně propracovaným KOS řadí především různé typy slovníků, jejichž struktura je většinou mělká a jejich hlavní funkcí je především odstraňování vágnosti jazyka. K nejpropracovanějším KOS, z opačné strany tohoto modelu, patří tezaury, ontologie a sémantické sítě, které mají velmi dobře propracovanou strukturu a pokrývají největší množství funkcí. Eliminují vágnost, zajišťují kontrolu synonym, poskytují hierarchické a asociativní vztahy, v případě ontologií a sémantických sítí navíc prezentují vlastnosti.

Various types of KOS

Zeng 2008 p.161



Obr. 3: Taxonomie KOS podle Zeng [2008]

Hodgeová [2000] seřadila KOS do tří skupin: rejstříky termínů (term lists), klasifikace a kategorie (classifications and categories) a rejstříky vztahů (relationship lists). K jednotlivým skupinám uvedla, jak již bylo řečeno, také příklady s jejich stručnou charakteristikou. Ta je založena na vlastnostech jako je struktura, komplexnost, vztahy mezi termíny a historická funkce.

1.1.1. Rejstříky termínů

Do skupiny *rejstříky termínů* řadí Hodgeová [2000] autoritní soubory (authority files), významové slovníky (glossaries), slovníky (dictionaries) a zeměpisné slovníky (gazetteers). Autoritní soubory jsou rejstříky termínů, které se používají ke kontrole různých podob jmen různých entit. Významové slovníky jsou většinou rejstříky termínů s definicemi, přičemž jednotlivé termíny mohou pocházet ze specifických oborů nebo určitých prací. Termíny jsou definovány v rámci určitého prostředí a zřídka obsahují i

jiné významy. Příkladem jsou Termíny životního prostředí (Terms of the Environment)⁵ Agentury pro ochranu životního prostředí (Environmental Protection Agency, EPA)⁶. Slovníky jsou abecední rejstříky slov a jejich definic. Slovníky mají na rozdíl od významových slovníků obecnější zaměření a poskytují přehled i více významů slov, jestliže se takové vyskytují. Slovníky také mohou obsahovat informace o původu slova, různých variantách pravopisu a vysvětlení i několika významů napříč různými disciplínami. Slovníky sice poskytují synonyma a díky definicím také příbuzné termíny, nicméně jednotlivá slova nejsou zařazena v rámci žádné jednoznačné hierarchické úrovně. Zeměpisné slovníky představují seznam jmen geografických míst. Tradiční zeměpisné slovníky byly publikovány jako knihy nebo se objevily ve formě rejstříku v atlasech. Každý termín může být identifikován jako řeka, město, škola apod. Příkladem jsou Americké kódy geografických jmen (U.S. Code of Geographic Name).

1.1.2. Klasifikace a kategorie

Do druhé skupiny *klasifikace a kategorie* patří předmětová hesla (subject headings), klasifikační schémata (classification schemes), taxonomie (taxonomies) a kategorizační schémata (categorization schemes). Předmětová hesla jsou typem systému, který se skládá z kontrolovaných hesel reprezentujících předmět jednotek v kolekci. Rejstříky předmětových hesel mohou být velmi rozsáhlé a pokrývat široké spektrum oborů. Přesto jsou struktury těchto rejstříků obecně velmi povrchní s limitovanou hierarchickou strukturou. Termíny klasifikační schéma, taxonomie a kategorizační schéma bývají často zaměňovány. Ačkoli příklad od příkladu mohou existovat drobné rozdíly, všechny tyto typy systémů organizace znalostí poskytují způsoby, jak jednotlivé entity rozdělovat do skupin nebo širších tématických úrovní. Tyto systémy také postrádají některé explicitní vztahy, které jsou vyjádřeny např. v tezaurech. Příklady klasifikačních schémat jsou Třídění Kongresové knihovny, Deweyho desetinné třídění nebo Mezinárodní desetinné třídění.

⁵ <http://www.epa.gov/OCEPATERMS/>

⁶ <http://www.epa.gov/aboutepa/index.html>

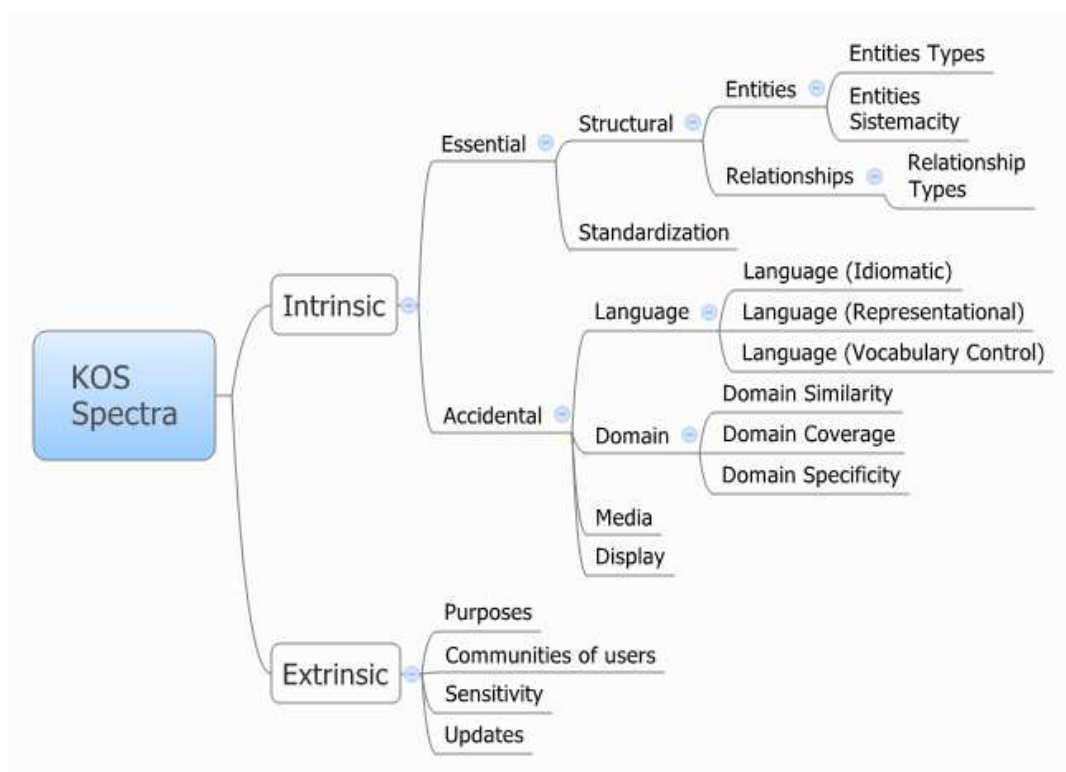
1.1.3. Rejstříky vztahů

Skupina *rejstříky vztahů* zahrnuje tezaury (thesauri), sémantické sítě (semantic networks) a ontologie (ontologies). Tezaury jsou založeny na konceptech a ukazují vztahy mezi jednotlivými hesly. Běžnou součástí tezurů jsou vztahy hierarchie, ekvivalence a asociace. Tyto vztahy jsou obecně označovány jako vztahy nadřazenosti a podřazenosti, synonymní vztahy a asociované nebo také příbuzné vztahy. Asociativní vztahy mohou být v některých systémech detailnější⁷. Tvorba a rozvoj tezurů jsou podpořeny jednak standardem pro jednojazyčné tezaury, jednak standardem pro tezaury vícejazyčné. Přestože se mnoho tezurů neřídí pravidly těchto standardů, jsou za tezaury považovány. Tezaury mohou obsahovat až padesát tisíc hesel. Řada z nich byla speciálně vyvinuta pro specifickou oblast či vědní disciplínu. S rozvojem procesů zpracování přirozeného jazyka souvisí také rozvoj sémantických sítí, v rámci kterých jsou definovány vztahy přesahující standardní, výše zmíněné, nadřazené, podřazené a příbuzné vztahy. Mohou zahrnovat specifické vztahy jako celek – část, příčina – následek nebo dítě – rodič. Nejvýznamnější sémantickou sítí je podle Hodgeové [2000] WordNet Princetonské univerzity. Ontologie mohou reprezentovat komplexní vztahy mezi objekty a zahrnovat pravidla a principy, které chybí v sémantických sítích. Ontologie popisující znalosti z určité oblasti jsou často propojeny se systémy data miningu a znalostního managementu.

1.2. Způsoby hodnocení systémů organizace znalostí

Stejně jako neexistuje shoda týkající se taxonomie KOS, neexistuje ani dohoda o tom, jak tyto systémy posuzovat, ani jaké by měly být parametry jejich hodnocení, podle kterých by KOS mohly být klasifikovány, i přesto, že je problematika KOS významné a často diskutované téma. Souza, Tudhope a Almeida [2010] navrhli model (viz obr. č. 4), ve kterém se pokusili jednotným a souvislým způsobem zachytit jednotlivé aspekty hodnocení KOS. Ty jsou rozděleny do dvou základních skupin: vnitřní parametry (Intrinsic) a vnější parametry (Extrinsic).

⁷ Na příklad UMLS (Unified Medical Language System) definoval více než 40 vztahů, z nichž je mnoho asociativních. UMLS je jeden z největších tezurů z oblasti medicíny a příbuzných věd.



Obr. 4: návrh parametrů hodnocení podle autorů Souzy, Tudhopea a Almeida [2010]

1.2.1. Vnitřní parametry hodnocení KOS

Vnitřní parametry hodnocení KOS souvisejí se samotnými KOS a jsou tak chápány jako parametry izolované od uživatelů a prostředí. Vnitřní parametry nebo také vlastnosti KOS se dále dělí na základní a náhodné. Základní vlastnosti KOS jsou úzce spojeny s typem, zatímco náhodné vlastnosti zahrnují charakteristiky daného příkladu KOS a mohou být proto různé pro každou implementaci stejného intelektuálního díla.

Základní vlastnosti KOS jsou rozděleny na strukturní vlastnosti a standardizaci. Strukturní vlastnosti se skládají z charakteristik entit a vztahů, které jsou představovány v KOS. Entity jsou ještě dále rozděleny na typy entit (např. slova, řetězce, čísla, koncepty atd.) a systematizaci entit (např. náhodná, abecední, systematická, enumerativní atd.). U vztahů jsou rozlišovány typy vztahů, které jsou vyjádřeny v systémech organizace znalostí (např. hierarchie, synonyma, asociativní vztahy atd.). Standardizace souvisí s existencí orgánů pro vytváření standardů, které mohou nastavit určitá pravidla nebo směrnice pro specifickou strukturu některého z KOS (např. ANSI/NISO Z39.19).

Náhodné vlastnosti jsou dále děleny na jazyk, doménu, média a zobrazení. V rámci kategorie *Jazyk* jsou rozlišeny jazyky na idiomatické (jednojazyčné, mnohojazyčné) a reprezentační (neformátovaný text, značkovací jazyky, formální jazyky), další vlastností je stupeň řízení (např. přirozený jazyk, umělý jazyk). Doména se dělí na podobnost, pokrytí a specifičnost. Podobnost odráží vztah entit k doméně. Pokrytí uvádí hloubku, ve které je doména pokryta (povrchní, hluboká), specifičnost souvisí se stupněm ontologie. Média souvisí s informačním podkladem (např. tištěný, digitální) a zobrazení informuje o způsobu prezentace informace (např. grafická, symbolická apod.).

1.2.2. Vnější parametry hodnocení KOS

Vnější parametry hodnocení KOS souvisejí s prostředím, ve kterém jsou KOS užívány. Zahrnují účel (jako indexace, klasifikace, reprezentace znalostí), komunitu uživatelů (knihovníci, weboví designéři, překladatelé), závislost a aktualizace.

1.3.Podpora výzkumu a vývoje v oblasti systémů organizace znalostí

1.3.1. Mezinárodní organizace ISKO

Významnou roli v podpoře výzkumu a vývoje v oblasti organizace znalostí hraje mezinárodní organizace *International Society for Knowledge Organization* (dále jen ISKO)⁸. Organizace ISKO byla založena v roce 1989 s cílem podporovat vývoj a aplikaci KOS a zároveň podporovat komunikaci a sdílení informací v rámci komunity členů ISKO, ale i mimo ni. V současné době má ISKO 400 členů (individuálních i institucionálních), odborníků z oblasti informační vědy, filozofie, lingvistiky, počítačové vědy a jiných vědních oborů. K nejvýznamnějším odborným aktivitám této organizace patří pořádání konferencí a to na mezinárodní (každé dva roky), ale i národní a regionální úrovni. Další významná činnost této organizace zahrnuje také vedení vědeckého časopisu *Knowledge Organization*. V jednotlivých zemích může ISKO působit prostřednictvím svých regionálních a národních poboček (např. britská pobočka

⁸ <http://www.isko.org/>

ISKO UK), které sdružují odborníky v dané oblasti. ISKO rovněž spolupracuje s dalšími národními a mezinárodními organizacemi na příklad s organizací Unesco⁹.

V letošním roce se mezinárodní konference ISKO konala v Londýně, ve Velké Británii ve spolupráci s univerzitou University College London. Konference, s podtitulem Facety organizace znalostí (*Facets of Knowledge Organization*), probíhala ve dnech 04. 07. 2011 – 05. 07. 2011 a byla věnována odkazu Briana C. Vickeryho (1918-2009), významného představitele informační vědy.

1.3.2. Sít' NKOS

Sít' *Networked Knowledge Organization Systems/Services* (NKOS)¹⁰ je komunita asi tří set odborníků z více než deseti zemí, kteří se zajímají o problematiku užití KOS v síťovém prostředí. Skupina odborníků NKOS pravidelně již od roku 1998 pořádá ve Spojených státech amerických a v Evropě semináře tématicky zaměřené na oblast KOS¹¹ [DCMI, 2009]. První seminář byl součástí konference *ACM Digital Library* konané v roce 1997, zaměřené na otázky tvorby interaktivních KOS na internetu a využití tezaurů v digitálních knihovnách. Poslední, v pořadí 9. evropský seminář, který se konal v září 2010 v Glasgow ve Skotsku se mimo jiné věnoval datovému modelu SKOS a oblasti linked data¹². Všechny konference a semináře, které se konají pod záštitou NKOS, jsou důležité události v procesu aplikace tezaurů v novém informačním prostředí World Wide Web [SHIRI; REVIE, 2000]. NKOS také ve spolupráci s OCLC zajišťuje správu elektronické e-mailové konference¹³.

Kromě již tradičních amerických a evropských seminářů pořádá NKOS ve spolupráci s dalšími organizacemi, jako je výše zmíněné ISKO, také různá specializovaná setkání. Poslední setkání proběhlo na letošní mezinárodní konferenci ISKO v Londýně, jejíž čtvrtá sekce byla uspořádána právě ve spolupráci s NKOS a věnovala se roli KOS

⁹ <http://www.unesco.org/new/en/unesco/>

¹⁰ <http://nkos.slis.kent.edu/>

¹¹ Přehled jednotlivých amerických a evropských seminářů včetně odkazů je umístěn na stránkách NKOS (<http://nkos.slis.kent.edu/>).

¹² Linked data „představuje publikační model pro zveřejňování strukturovaných dat na webu. Je vymezen principy a doporučeními, která stanovují, jakým způsobem mají být data na web publikována“ [MYNARZ; ZEMÁNEK, 2010].

¹³ <http://listserv.oclc.org/archives/nkos-l.html>

v informačních vyhledávacích aplikacích. V rámci této sekce vystoupila mimo jiné i Vivien Petras, která hovořila o současných problémech a výzvách, které doprovázejí užití KOS nejen v procesech vyhledávání informací, ale i v prostředí sémantického webu. Za nejpopulárnější KOS v rámci svého vystoupení označila Wikipedii, MeSH a WordNet.

2. VLIV WEBOVÉHO PROSTŘEDÍ NA SYSTÉMY ORGANIZACE ZNALOSTÍ

KOS jsou tradičně vytvářeny a užívány v knihovnách po celém světě jako nástroje určené k organizaci a vyhledávání dokumentů. *„Organizace a reprezentace znalostí tradičně představuje tvorbu katalogů a indexů pro dokumenty jakéhokoliv druhu. Zahrnuje popis vlastností dokumentu a reprezentuje jeho intelektuální obsah. Knihovny po celém světě mají za sebou dlouhou historii v zaznamenávání těchto informací o dokumentech. Indexy a knihovní katalogy byly a jsou vytvářeny proto, aby pohodlně pomáhaly uživatelům nalézt a lokalizovat dokument. Taková praxe je datována až několik tisíc let zpátky“* [QIN, 2000]. V době svého vzniku a ještě dlouho poté byly KOS izolovány ve svém mateřském prostředí. S nástupem počítačů, vznikem Internetu a World Wide Web a dalším rozvojem informačních a počítačových technologií se však tvůrcům KOS otevřely nové možnosti pro jejich využití. Ačkoli již bylo mnoho KOS převedeno do elektronické formy, a dokonce zpřístupněno online, Johnson [2004] konstatuje, že pro vyhledávání zatím tyto nástroje příliš užitečné nejsou. Klíčem k úspěchu je totiž jejich aplikace v rámci sémantických webových technologií a standardů zaručujících jejich efektivitu v procesu vyhledávání informací. Užití sémantických webových technologií je mimo jiné také předpokladem k naplnění vize Tima Bernerse-Lea¹⁴ o sémantickém webu.

2.1. Charakteristika webu v kontextu organizace informací

World Wide Web (označovaný také jako "WWW" nebo "web") je *„distribuovaný hypertextový internetový informační systém, v němž dokumenty obsahují odkazy na jiné místní nebo vzdálené dokumenty. Informační nástroj využívající text, grafiku, zvuk, video. Webové dokumenty jsou vytvořeny nejčastěji v jazyku HTML, pro přenos sítí se používají protokoly HTTP, FTP a další. Klientský program používaný pro zobrazování webových stránek a pro pohyb v informačním prostoru WWW se nazývá prohlížeč* [SKLENÁK, 2003].

Zjednodušeně řečeno, web se vyvinul do podoby, jak jej známe dnes, protože je jednoduché vytvořit webovou stránku a tu prolinkovat s ostatními. Neexistuje ani limit

¹⁴ Tim Berners-Lee je vynálezce World Wide Webu a ředitel konsorcia W3C (<http://www.w3.org/>).

pro počet webových stránek, ani pro počet linků. V průběhu posledních let roste web exponenciálně [CHOWDHURY G. G.; CHOWDHURY S., 2007, s. 132]. Tvorba webových stránek navíc nevyžaduje žádnou koordinaci, ani centrální autoritu a zahrnuje extrémně nízké, ne-li nulové náklady [MILES, 2008]. Taková praxe je velmi odlišná od té, kterou známe z bibliografického informačního prostředí, ve kterém organizace bibliografických informačních zdrojů zahrnující katalogizaci, klasifikaci a předmětovou indexaci byly vždy procesy závislé na člověku, procesy náročné na peníze a čas. Aby knihovny snížily náklady a ušetřily čas spojený s těmito činnostmi, využívají centralizované zpracování informací a sdílení zdrojů, které je v kontextu roztržité povahy webu nepředstavitelné. *„Přestože je stále ve větší míře vyžadován snadný přístup a organizace webových informačních zdrojů, faktem zůstává, že k jejich organizaci chybí někdo nebo něco, co by zajistilo financování takové činnosti“* [CHOWDHURY G. G.; CHOWDHURY S., 2007].

S růstem webu se identifikace, organizace a vyhledávání informací stává více komplexním procesem. Informační soubory na webu se mohou vyskytovat v různých podobách, od prostého textu až k digitálnímu videu. Každý soubor se navíc může vyskytovat v několika různých formátech. Web tak představuje obrovskou kolekci neorganizovaných informačních zdrojů, ve kterém se, v ostrém kontrastu s tradičními knihovnami, které pracují s relativně statickými informacemi, příliš statických a fyzických materiálů nevyskytuje. S tím souvisí jedna z typických charakteristik webu, kterou je vysoká frekvence změn. Webové stránky se mění poměrně často a to ve smyslu lokace, vlastnictví, struktury, obsahu apod. Něco takového je v prostředí knihovny nepředstavitelné, alespoň pokud jde o klasický knižní fond, neboť stane-li se jednou kniha součástí knihovní sbírky, jen těžko mění vlastníka, nebo dokonce obsah.

Web a informace založené na HTML formátu zvýšili přístup k online informačním zdrojům, zároveň však limitují způsoby, jakými mohou být tyto zdroje využity. Webové založené indexování a vyhledávače zpřístupňují plné texty miliónů online dokumentů, na straně druhé k nim ztěžují přístup v jakékoli organizované, systematické podobě. Hlavním zdrojem problému je, že web reprezentuje tematickou a nebo logickou organizaci použitím stejných nástrojů, které užívá k reprezentaci skutečných cílových informací (HTML dokumenty). To znamená, že katalogy a indexy jsou ztraceny v mixu všech webových stránek. Johnson [2004] použil pro lepší představu této situace analogii

k tradiční knihovně tištěných knih. Podle něj si lze představit knihovnu, ve které se nacházejí pouze hromady knih místo katalogu a knihy jako samostatné entity. Různé díly knih jsou zde nahodile rozptýlené v hromadě knih, nesnadno rozeznatelné od zbytku. Katalogy reprezentují pouze dílčí části celku, a proto není jisté, jestli všechny knihy v kolekci byly katalogizovány. V takové knihovně neexistuje nikdo, kdo by mohl tvrdit, že zná lokaci všech katalogů. Uživatel možná ví, podle své vlastní zkušenosti, o určitých dílech katalogů, o které se zajímá, ale musel by se spolehnout na ostatní lidi, aby se dozvěděl, kde se nacházejí jiné druhy katalogů a indexů a to vše s vědomím, že se tam někde v prostoru možná nacházejí i další, tedy takové které by snad mohly být užitečnější, ale o kterých neví. Pořadí knih na regále je nahodilé, bez jakéhokoli tematického zařazení. Když uživatel jednou nalezne jeden nebo více katalogů, které uspokojí jeho informační potřebu, má tendenci vracet se k nim, přestože nutně nezahrnují nejrelevantnější nebo aktuální informace. Uživatelé si vybírají informační zdroje podle toho, jestli je znají a podle snadnosti jejich užití spíše než podle kvality informací, které zahrnují.

Tato analogie výstižně charakterizuje nejen podobu webu, ale především chování uživatelů, kteří prohledávají web tak, že se zaměřují na to, co mohou nalézt prostřednictvím několika oblíbených portálů nebo vyhledávačů, zatímco jim s velkou pravděpodobností unikají relevantnější informace. V jiných případech se zase spoléhají na příklad na své přátele či kolegy, kteří jim pomohou lokalizovat relevantní a užitečné portály a webové stránky, místo toho, aby by byli schopni nalézt je přímo [JOHNSON, 2004].

Pokud jde o organizaci informací a informační management, pak se nejsilnější stránka webu může proměnit také v jeho největší slabost. Idea webu – globálně prolinkované kolekci dokumentů – předem brání organizačním principům, podle kterých by mohl uživatel začít prohlížení kolekce. Podle Johnsona [2004] jistý druh organizace představuje Domain Name Service (dále jen DNS), ale pouze z pohledu serverových lokací a nikoli z pohledu tematické organizace. Pro ty, kteří se chtějí v prostředí webu orientovat pomocí tématu nebo předmětu, DNS poskytuje pouze malé využití. Znalost toho, že nějaká webová stránka je pod "org" a ne pod "com", nenabízí uživateli žádnou pomoc. Relevantní dokumenty se totiž mohou nacházet pod jednou, druhou, ale také pod žádnou z těchto domén.

K informačním zdrojům, jejichž lokace není známá, je možné přistupovat prostřednictvím vyhledávacích služeb, které lze z funkčního hlediska rozdělit do dvou hlavních skupin. První skupinu tvoří vyhledávací stroje označované také jako vyhledávače. Do druhé skupiny patří předmětové katalogy. Zatímco vyhledávače „indexují slova nebo termíny, které se vyskytují ve WWW dokumentech“, předmětové katalogy „klasifikují dokumenty nebo celé servery podle dané předmětové klasifikace nebo taxonomie“ [SKLENÁK, 2001]. Předem vytvořené kategorie předmětových katalogů se tak mohou řadit k zástupcům KOS. Každá z uvedených metod přístupu k informacím má v procesu vyhledávání informací své výhody a nevýhody, které byly předmětem mnoha studií. Faktem ale je, že se různé vyhledávače staly v poslední době velmi populární.

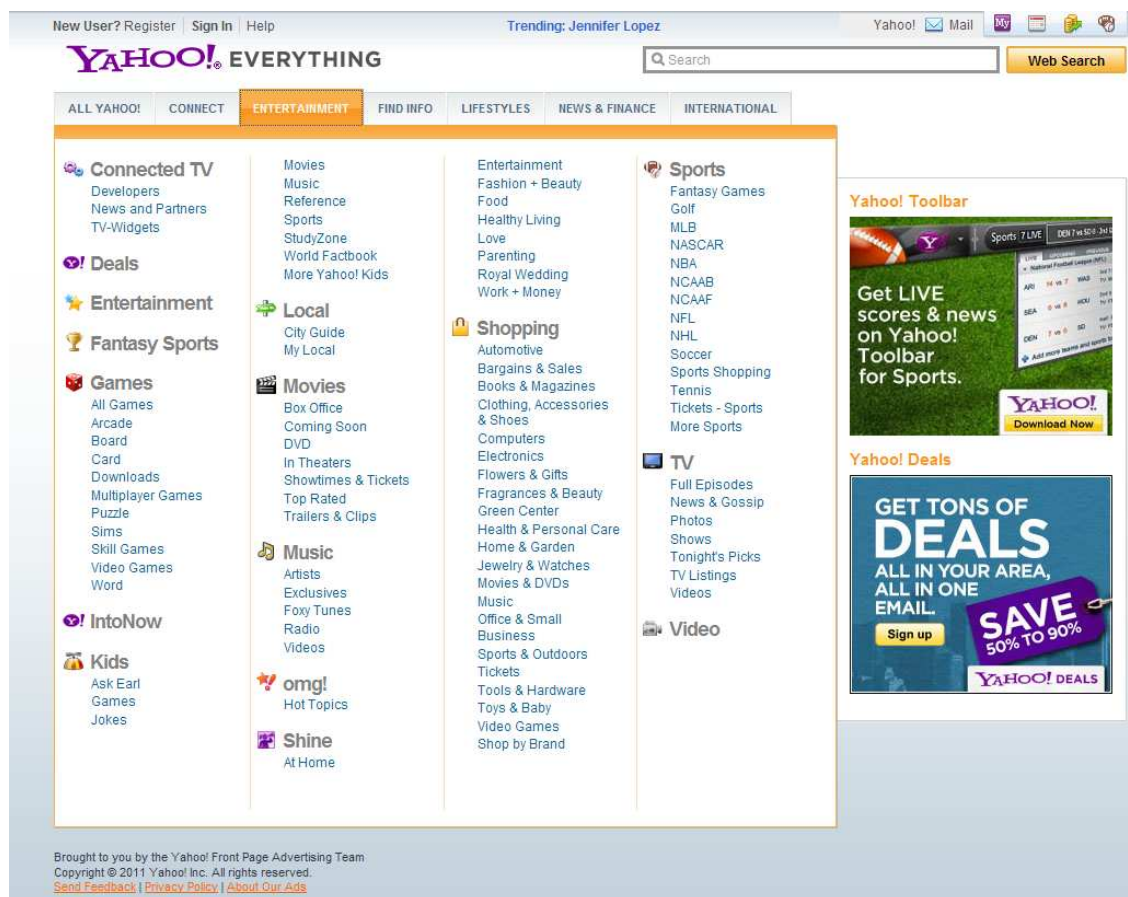
Vyhledávače

Přístup k informacím prostřednictvím vyhledávače má všechny výhody, ale i slabosti klasického online vyhledávání informací a to přesto, že jsou vyhledávací techniky a algoritmy vylepšovány každý den. Jakkoli se vyhledávačům nedá odepřít jejich vysoká popularita, má tento způsob vyhledávání několik limitů. Miles [2005] hovoří o problémech spojených s webovými zdroji, které nemají žádný nebo pouze malý textový obsah (na příklad obrázky, zvuk, video apod.). Dále se zabývá otázkou přesnosti výsledků, které jsou podle něj nízké, jestliže mají termíny přirozeného jazyka aplikované do procesu vyhledávání mnoho významů (na příklad v angličtině slova jako bank, watch, chip apod.). Na obecné úrovni se tedy jedná o problémy plynoucí z přirozenosti jazyka – jevy jako je synonymie, kvazisynonymie apod. Podle Milese [2005] rovněž neexistuje základna pro pomoc či asistenci uživatelům při vyhledávání jako je rozšíření dotazu, zpřesňování výsledků vyhledávání nebo překládání, ani možnost vyhledávání napříč přirozenými jazyky nebo rámcové vyhledávání ve smyslu symbolického jazyka.

Předmětové katalogy

Předmětové katalogy (také adresáře nebo klasifikační stromy) mají předem vytvořené kategorie, do kterých jsou poté mapovány webové stránky. Takový přístup je zdrojově náročný, neboť zahrnuje nejen tvorbu organizační struktury, ale i mapování webových stránek do vytvořené struktury podle jejich obsahu. Předmětové katalogy jsou často součástí webových portálů. Webový portál je možné charakterizovat jako „webové

sídlo, které poskytuje širokou škálu služeb a informací, často s možností jejich přizpůsobení uživateli podle osobních potřeb a zájmů. Nabídka zahrnuje především vyhledávání webových informačních zdrojů a dokumentů, freemailovou službu, denní zpravodajství, mapy, vyhledávání kontaktních informací na osoby i firmy, zábavu (hry, horoskopy, vtipy apod.), elektronické obchody, bezplatný prostor pro publikování webových stránek apod. Naprostá většina nabízených služeb je pro koncového uživatele bezplatná díky ziskům z reklamy. Široce (obecně) zaměřené portály se označují jako horizontální portály, na rozdíl od vertikálních portálů“ [CELBOVÁ, 2003]. Přestože dnes mají uživatelé k dispozici široké spektrum webových portálů, ze kterých si mohou vybrat, často mají obtíže nalézt nejrelevantnější z nich, tedy takové, které by odpovídaly jejich informační potřebě [JOHNSON, 2004]. Příkladem webového portálu je Yahoo!, který svým uživatelům, kromě dalších služeb, nabízí speciálně organizované navigační hierarchie linků k webovým stránkám (na příklad viz obr. č. 5). Podle Aitchinsona a Clarka [2004] popularita Yahoo! prokázala, že netrénovaní uživatelé jsou spokojení s prohlížením jednoduše organizovaného předmětového katalogu raději než s aktivní tvorbou termínů pro vyhledávání. Využívání takových hierarchií však v sobě může skrývat jisté nebezpečí. V těchto případech totiž o tom, co je relevantní, rozhodují tvůrci portálů a nikoli uživatelé. Jejich tvorba může být navíc ovlivněna sponzorstvím nějakého komerčního subjektu či dokonce politické strany.



Obr. 5: Zobrazení kategorií (záložka "zábava") nabízených prostřednictvím webového portálu Yahoo!

2.2. Před webová éra KOS

První záznam o historii tezaurů se objevuje v díle Alana Gilchrista *Thesaurus in Retrieval* publikovaném v roce 1971. První tezaury byly vyvinuty v papírové éře. Vyhledávání v papírových kartotékách bylo většinou vázáno k jednomu místu, ke kterým informační profesionálové zprostředkovávali uživatelům přístup. Personál byl školen, aby uměl efektivně užívat tento nástroj pro vyhledávání. V 70. a 80. letech došlo k rozšíření elektronických databází buď prostřednictvím online hostitelů nebo později na discích CD-ROM. Stále byl však přístup přes jeden terminál s kolekcí tištěných tezaurů kolem. Školení uživatelé byli schopni zadávat příkazy a využívat k nim kontrolované slovníky. Kvalita kontroly jak na vstupu, tak výstupu zajistila efektivnost výsledku vyhledávání.

Zeng a Chan [2004] považují za nejvýznamnějším období historického vývoje KOS etapu od počátku šedesátých let do konce osmdesátých let. Pro tuto dobu je příznačný

růst online databází, a proto toto období označují jako online éru. První online databáze byly zejména indexační a abstraktové databáze nebo sekundární zdroje. Důležitost kontrolovaných slovníků v procesu zvyšování efektivity vyhledávání informací (měřeno v podstatě na základě přesnosti a úplnosti výsledků) bylo potvrzeno komplexními experimenty v oblasti vyhledávání informací, jakým byl Cranfieldský test a výzkum MEDLARS.

2.3. Webová éra KOS

Vynález a podoba webu velmi ovlivnily aplikaci KOS. Před jeho vznikem byly KOS účinnými nástroji v kontrolovaném prostředí s jasně definovanými strukturami a způsoby pro přístup k informacím. Na webu, kde dochází ke kombinaci heterogenních formátů a struktur s velkým množstvím zdrojů a obsahů, je téměř nemožné aplikovat tento pracovní model.

Shiri; Revie [2000] uvádí, že jsme svědky rostoucího entusiasmů mezi tvůrci tezaurů, kteří chtějí tyto nástroje poskytovat na web pro různé využití. Důvod takového nadšení a rostoucí dostupnost online tezaurů jsou úzce spojeny se vznikem webu. Konkrétně tyto změny souvisí s obrovským nárůstem informačních zdrojů vyžadujících lepší předmětovou identifikaci, s migrací tradičních informačních zdrojů na web, s naléhavou potřebou popisu zdrojů s využitím již existujících KOS, dále jsou tyto změny spojeny s problémem kvality nestrukturovaných informací na webu a s potřebou poskytnout uživatelům tezaury pro rychlý a snadný přístup k lépe organizovaným informacím. Webovou éru tak charakterizuje určitý paradox. Ačkoli je zdánlivě potřeba a příležitost aplikovat tezaury v procesu vyhledávání informací větší než kdy předtím, sami uživatelé odolávají tomu, aby je aplikovali, a to i přesto, že se některé KOS adaptovaly do digitálního prostředí, v němž je však jejich efektivní využití oproti snaze stále malé. Mezi jinými je to také z těchto důvodů:

- Limitovaný vývoj a vymezení konceptuálně orientovaných tezaurů. Aplikace lexikálních tezaurů v dynamickém prostředí jako je web neposkytuje adekvátní výsledky ve srovnání se snahou věnovanou indexačnímu procesu.
- Počáteční absence vhodných standardů a modelů k reprezentaci např. XML (Extensible Markup Language), RDF (Resource Description Framework) nebo SKOS (Simple Knowledge Organization System).

- Rychlý úspěch vyhledávačů, které prokázaly svou efektivnost využitím nejrozumnějších technologií, které je v mnoha případech těžké kritizovat.
- Představa tezaurů jako zastaralého nástroje nebo nástroje s limitovaným užitím [PASTOR-SANCHEZ; MARTINEZ MENDEZ; RODRÍGUEZ-MUÑOZ, 2009].

K překonání výše zmíněných překážek, ke zvýšení efektivity KOS a k rozvoji konceptu sémantického webu¹⁵ mohou knihovny přispět tvorbou a rozvojem různých slovníků, tezaurů, klasifikačních schémat, tedy obecně KOS, a jejich implementací do formátů umožňujících vnoření webových technologií. Nejednou Eric Miller [2004] ve svých prezentacích vyzval komunitu knihovníků, aby se aktivně zapojila do vývoje sémantického webu a to především v oblastech:

- zpřístupňování kolekcí s využitím sémantických webových technologií,
- webifikace tezaurů [MILLER, 2004].

Z hlediska předmětu této práce je nejzajímavějším bodem právě webifikace tezaurů. Obecně může webifikace¹⁶ znamenat „převod souborů jako jsou text nebo obrázky z jejich původních formátů do obsahu, který může být zobrazen na webu“ [WEBIFY, 2005-2009]. Proces webifikace je v tomto kontextu chápán jako převod existujících KOS do sémantických webových standardů jako jsou na příklad Simple Knowledge Organization System (SKOS)¹⁷ nebo Web Ontology Language (OWL)¹⁸ zahrnující užití jednotných identifikátorů Uniform Resource identifier (URI), jazyků Resource Description Framework (RDF).

Tezaury, ale i jiné typy KOS, které již byly na webu publikovány a zpřístupněny se mohou lišit v nástrojích nebo technikách, které byly za tímto účelem použity. Shiri;

¹⁵ Sémantický web je „koncept webu založený na obsahu, který je vytvořen a strukturován podle určitých pravidel a standardů a umožňuje tak efektivnější a snadnější vyhledávání informací. Realizace sémantického webu předpokládá implementaci standardů pro sémantickou (RDF), strukturální (XML) a syntaktickou (URI) složku architektury webových dokumentů; výsledkem aplikace uvedených standardů bude konzistentní logická struktura dat, která bude implicitně vyjadřovat význam zaznamenaných informací“ [CELBOVÁ, 2003a].

¹⁶ Termín "webifikace" je převzat z anglického slova *webifying*.

¹⁷ <http://www.w3.org/2004/02/skos/>

¹⁸ <http://www.w3.org/TR/owl-features/>

Revie [2000] rozčlenili webové založené tezaury do kategorií podle použitých publikačních formátů a struktur. Jedná se o tyto skupiny tezurů:

- tezaury v jednoduchých textových statických formátech,
- tezaury ve statickém formátu HTML (bez efektivního užití hyperlinků),
- tezaury v dynamickém formátu HTML (užití hyperlinků),
- tezaury s pokročilými vizuálními a grafickými rozhraními,
- tezaury v XML formátu.

Webové založené tezaury mohou být také rozděleny do dvou obecných kategorií podle jejich funkčnosti a použití. Do první skupiny se řadí samostatně stojící tezaury, které nejsou součástí žádného informačního systému. Do té druhé patří naopak ty tezaury, které jsou plně integrovány do databáze nebo informačního systému (např. thesaurus ERIC¹⁹).

Navzdory některým počátečním obavám jsou knihovníci v post-webové éře považováni za poskytovatele služeb v oblasti organizace znalostí, což je do velké míry dáno jejich zkušenostmi s touto problematikou. Někteří dokonce tvrdí, že díky růstu digitálních knihoven a digitálního světa obecně, se lidská klasifikace stane ještě významnější. A to pravděpodobně z toho důvodu, že počítače nebudou možná nikdy schopny opravdově zachytit drobné sémantické nuance jazyka [SAMURE; SHIRI, 2008].

¹⁹ <http://www.eric.ed.gov/ERICWebPortal/thesaurus/thesaurus.jsp>

3. VYBRANÉ TRENDY SYSTÉMŮ ORGANIZACE ZNALOSTÍ

Trendy v oblasti KOS jsou neodmyslitelně spjaty s vlivem webového prostředí na tyto systémy, ale souvisí také s rozvojem informačních a počítačových technologií, které umožňují nahrazovat často velmi časově a finančně náročnou intelektuální činnost spojenou s tvorbou, údržbou a aplikací KOS. V zahraniční literatuře se otázce trendů v oblasti organizace znalostí věnovalo několik autorů [AITCHISON; CLARKE, 2004; SAUMURE; SHIRI, 2008; ZENG; CHAN, 2004].

Samure; Shiri [2008] se ve své práci zabývali trendy v před- a v post- webové éře. Na základě provedené kvalitativní analýzy článků z databáze Library, Information Science, and Technology Abstracts (dále jen LISTA) identifikovali a stručně charakterizovali dvanáct široce zaměřených trendů. Konkrétně se jednalo o analýzu článků, pro jejichž vyhledání bylo použito termínů *organizace znalostí* (knowledge organization) a nebo *organizace informací* (information organization) v titulcích, abstraktech nebo popisech. Trendy byly identifikovány v těchto kategoriích:

- organizace obchodních a korporativních informací,
- organizace znalostí s využitím strojů,
- knihovníci jako organizátoři znalostí na webu,
- interoperabilita,
- katalogizace a klasifikace,
- klasifikace webových informací,
- digitální ochrana a digitální knihovny,
- aplikace metadat a jejich užití,
- poznávání,
- vzdělávání,
- indexování a abstraktování,
- iniciativy tezaurů.

Studie Samure; Shiri [2008] je přínosná, jelikož odkrývá trendy v oblasti organizace znalostí, která v posledních čtyřiceti letech prošla značnou proměnou. Současně je však

nutné si uvědomit dva limitující faktory této práce, které přiznávají i sami autoři. Prvním z nich je fakt, že studie analyzuje pouze články, které se objevily v databázi LISTA. Druhým omezením je použitá terminologie, která skrývá riziko nedorozumění v různých vědních oborech. Existují vědci pohybující se mimo obor informační vědy a knihovnictví, kteří používají termín *organizace znalostí* bez dřívější zkušenosti ve vztahu k dlouhodobé tradici tezurů či klasifikačních schémat. Na příklad počítačové vědy, konkrétně experti na umělou inteligenci, používají termín *organizace znalostí* a *reprezentace*, ale s rozdílnými konotacemi a významy [SAUMURE; SHIRI, 2008].

Zatímco Samure; Shiri [2008] identifikovali dvanáct "trendových" kategorií, Aitchinson a Clarke [2004] uvedli dva hlavní trendy pro oblast tezurů. Ty odvodili na základě konfrontace s těmito výzvami:

- Přístup k informacím prostřednictvím různých portálů, bran a vyhledávačů, z nichž je mnoho přizpůsobeno určitému publiku a předmětovým oblastem. Neexistuje univerzální tezaurus, ale velké množství různých slovníků pro různé aplikace.
- V prostředí "jednou publikuj, využij několikrát" je velmi těžké předpovídat, ve kterém systému nebo síti se požadovaný dokument nakonec vyskytne.
- S obrovským množstvím autorů, webmasterů, systémových administrátorů apod. nemůže být kontrola kvality vynucená kvůli organizačním překážkám.
- Otázka toho, jak naučit koncové uživatele správně používat tezaurus. Zkušenost většiny poskytovatelů informací říká, že uživatel nechce být zatěžován ničím komplikovaným, na tezaurus je však tak nahlíženo [AITCHISON; CLARKE, 2004].

Střet s těmito výzvami, jak již bylo řečeno, vedl ke dvěma hlavním trendům, které lze ve vývoji tezurů pozorovat. Prvním trendem je adaptace KOS, od níž si autoři slibují rychlost, jednoduchost a intuitivnost KOS v rámci jejich užití. Druhý trend představuje potřebu interoperability KOS.

Aitchinson a Clarke [2004] vztahují interoperabilitu ke KOS a chápou ji jako způsob tvorby těchto systémů tak, aby byly schopné snadné integrace nejen s jinými KOS, ale také s populárními aplikacemi jako jsou na příklad systémy správy obsahů, tagovací rozhraní, vyhledávače a portály. V pojetí Samure; Shiri [2008] však není

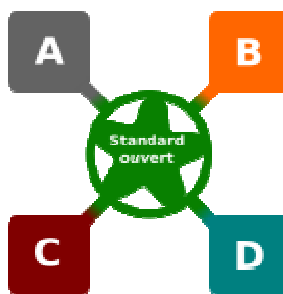
interoperabilita spjata s KOS, ale je spojena se sdílením digitálních zdrojů v rámci různých organizací. Zahrnuje tak otázku užití metadatových standardů a schémat umožňující výměnu dat a informací napříč repozitáři.

Jiným příkladem je práce Zeng; Chan [2004], kteří hovoří o dvou základních aspektech KOS, na které narážíme v internetovém prostředí, a na které je při tvorbě a rozvoji KOS nutné brát ohled. Prvním neopomenutelným aspektem je síťově propojené prostředí a druhým jsou digitální formáty. A právě síťově propojené prostředí velmi ovlivnilo aplikaci KOS a umocnilo potřebu interoperability těchto systémů.

S ohledem na různorodost a množství KOS, podrobněji popsanych v první kapitole, není možné v této práci zachytit a zabývat se všemi trendy, které se této oblasti dotýkají. Proto je tato kapitola věnována pouze vybraným trendům, které považuji za stěžejní a zároveň klíčové ve vztahu k Polytematickému strukturovanému hesláři (PSH) a Předmětovým heslům Kongresové knihovny (LCSH), tedy k tradičním představitelům KOS. *„Nadšení pro užití moderních webových technologií k publikaci tezaurů na webu vedlo k růstu jejich počtu a k potřebě přemýšlení o způsobech jejich interoperability, stejně jako vedlo k potřebě přístupu a užití různých tezaurů v procesu vyhledávání“* [SHIRI; REVIE, 2000]. Na základě prostudované literatury a vzhledem ke kontextu PSH a LCSH se budu zabývat trendy v oblasti interoperability a aplikace webových technologií.

3.1. Interoperabilita

Interoperabilita je velmi široký pojem, který se může vztahovat k mnoha oborům a oblastem lidské činnosti. Známá je na příklad interoperabilita na politické úrovni (např. vojenská spolupráce), na úrovni železniční infrastruktury, dále se tento pojem používá v souvislosti s technickou interoperabilitou různých přístrojů, softwarů apod. Obecně může být interoperabilita definována jako *„součinnost, schopnost spolupracovat, univerzálnost. Např. schopnost aplikačního programu pracovat v prostředí heterogenních sítí nebo vyměňovat si data s jinou aplikací“* [TKAČÍKOVÁ, 2000]. Jiná definice charakterizuje interoperabilitu jako *„schopnost produktu nebo systému spolupracovat s jinými produkty a systémy bez jakéhokoli omezení v přístupu a implementaci“* (viz obr. č. 6) [GROUPE, 2009].



Obr. 6: Grafické znázornění interoperability
[GROUPE, 2009]

Interoperabilita, která se vztahuje k předmětu této práce představuje vlastnost, která umožňuje, aby KOS byly schopné vzájemného propojení či integrace nejen s jinými KOS, ale také s jinými aplikacemi. Jak již bylo naznačeno, potřeba interoperability je zmiňována v dílech mnoha autorů [AITCHISON; CLARKE, 2004; SAUMURE; SHIRI, 2008; SHIRI; REVIE, 2000; ZENG; CHAN, 2004]. „V současném síťově propojeném prostředí je interoperabilita KOS nevyhnutelnou otázkou“ [ZENG; CHAN, 2004].

Ačkoli by se mohlo zdát, že je koncept interoperability KOS novým fenoménem, není tomu tak. A to i přesto, že Samure; Shiri [2008] konstatují, že se termín *interoperabilita* v před-webové éře v jimi analyzovaných záznamech neobjevuje, zatímco později, po vynálezu webu, je použití tohoto termínu všudypřítomné. Ještě dlouho před elektronickou érou knihovníci a informační profesionálové aplikovali řadu metod, jejichž cílem byla redukce rozdílů mezi různými typy KOS užívaných v rámci jednoho systému. Rozdíl v dnešní době spočívá v tom, že zatímco v minulosti byly tyto metody téměř výlučně založeny na práci člověka spočívající v náročné intelektuální činnosti, dnešní projekty k tomuto účelu využívají moderní počítačové a technologicky vyspělé metody. Jakkoli jsou však tyto metody stále častěji využívány, podle Zeng; Chan [2004] bude mapování založené na intelektuální snaze člověka ještě nějakou dobu koexistovat s technologiemi počítačového mapování.

Během posledních let proběhlo několik projektů, které si kladly za cíl nějakým způsobem sjednotit KOS v jeden integrovaný zdroj nebo v jedno uživatelské rozhraní. V těchto projektech byla pozornost věnována především tvorbě interoperability mezi KOS různých jazyků a nebo struktur. Výsledky vedly k integraci, mapování a tvorbě takových KOS, které jsou vhodné pro sdílení informací v síťovém prostředí. Tyto projekty se liší jak v cílech mapování, tak v metodách, které k tomuto účelu byly

použity. Jedním z projektů, které řešily jazykovou interoperabilitu, je na příklad *MACS* (*Multilingual Access to Subjects*), jehož cílem je propojení kontrolovaných slovníků ve třech jazycích (angličtině, francouzštině a němčině). Provedení mezi systémové interoperability reprezentují projekt *WebDewey*²⁰ mapující Předmětová hesla Kongresové knihovny (kontrolovaný slovník) s Deweyho desetinným tříděním (klasifikační schéma) a projekt Kongresové knihovny *Classification web*²¹ (klasifikační web) (viz dále).

Základním problémem ve formování interoperability mezi různými existujícími KOS jsou právě jejich rozdílné struktury a vlastnosti. Na příklad kontrolované slovníky typicky využívají k organizaci abecedu, zatímco klasifikační schémata pracují s notací. Jednotlivé vztahy hierarchie, asociace a ekvivalence se v obou typech KOS mohou navíc objevovat v různých makro strukturách. Jak tezaury, tak klasifikační schémata jsou nástroje sloužící k organizaci znalostí a jako takové se vzájemně doplňují, což je jeden z důvodů, proč jsou tyto systémy vzájemně mapovány nebo linkovány.

3.1.1. ISO: 25964

Významnou aktivitu v oblasti interoperability KOS představuje nově vznikající mezinárodní standard ISO 25964: *Information and Documentation – Thesauri and interoperability with other vocabularies*. Tento standard bude mít dvě části, které aktualizují, upravují a nahrazují existující mezinárodní standardy ISO 2788 pro jednojazyčné tezaury z roku 1986 a ISO 5964 pro vícejazyčné tezaury z roku 1985, stejně jako nahradí BS 8723 Structured vocabularies for information retrieval - Guide²². Klíčovým rozdílem mezi nově připravovaným standardem ISO 25964 a jeho předchůdci je zaměření na interoperabilitu. Clarke [2011] charakterizuje tento rozdíl jako proměnu samostatně stojících papírových tezurů v tezaury užitečné v síťovém propojeném prostředí. Vývoj standardu zajišťuje pracovní skupina WG8, která byla ustavena v rámci subkomise ISO TC46/SC9. Ke členům této skupiny pocházejících až z patnácti různých zemí světa (např. Švédsko, Francie, Německo, Belgie, Bulharsko, Španělsko

²⁰ <http://www.oclc.org/dewey/versions/webdewey/default.htm>

²¹ <http://classificationweb.net/>

²² Jedná se o britský standard publikovaný mezi léty 2005-2008 v pěti oddělených částech (1. část: definice, symboly a zkratky, 2. část: tezaury, 3. část: slovníky jiné než tezaury, 4. část: interoperabilita mezi slovníky, 5. část: interoperabilita mezi slovníky a jinými komponenty rešeršních systémů).

atd.) patří na příklad Johan De Smedt²³, Stella Dextre Clarke²⁴, Alan Gilchrist²⁵, Traugott Koch²⁶, Marianne Lykke²⁷, Douglas Tudhoupe²⁸, Leonard Will²⁹ nebo Marcia Zeng³⁰.

První část standardu ISO 25964 (*Thesauri for information retrieval*) by měla být zveřejněna ještě v roce 2011 a bude se dotýkat vývoje a ustavení tezaurů (jednojazyčných i vícejazyčných). Konkrétně bude pokrývat tyto oblasti:

- obsah a konstrukce tezaurů (jednojazyčných i vícejazyčných),
- aplikace facetové analýzy na tezaurus,
- návod na vývoj a údržbu tezaurů,
- funkční požadavky na softwaru pro údržbu tezaurů,
- datový model a derivované XML schéma (datový model ISO 25964 je z velké části kompatibilní se standardem SKOS).

Druhá část standardu ISO 25964 (*Interoperability with other vocabularies*), jejíž zveřejnění je připravováno na rok 2012, je zaměřená na interoperabilitu tezaurů a jiných typů kontrolovaných slovníků (klasifikační schémata, taxonomie, předmětová hesla, ontologie, autoritní soubory aj.). Clarke [2011] konstatuje, že snahou druhé části není standardizovat tyto různé typy kontrolovaných slovníků. Ty budou popsány pouze do té míry, aby mohly být použity pro interoperabilitu s tezaury, zejména půjde o to, jakým způsobem implementovat mezi těmito systémy mapování. V pracovní verzi druhé části standardu (ISO 25964-2) je mapování definováno jako: „*vztah mezi koncepty v jednom*

²³ Expert na reprezentaci znalostí a modelování informací, pracuje pro společnost TenForce, která se mimo jiné specializuje na sémantické webové technologie.

²⁴ Předsedkyně britské pobočky ISKO UK, specialista v oblasti aplikace znalostních struktur do procesů vyhledávání informací.

²⁵ Zabývá se oblastí informačního managementu, zejména se pak věnuje problematice organizace znalostí.

²⁶ Vědecký pracovník digitální knihovny *Max Planck Digital Library* v Německu.

²⁷ Profesorka informační architektury a interakce na Aalborgské univerzitě v Dánsku.

²⁸ Profesor Glamorganské univerzity ve Velké Británii, editor časopisu *The New Review of Hypermedia and Multimedia* a člen NKOS.

²⁹ Správce financí ISKO UK.

³⁰ Profesorka na Kentské státní univerzitě v Ohio, k jejím zájmům patří na příklad organizace a reprezentace znalostí.

slovníku a jednom nebo více konceptů v jiných“. Mapování je tak proces, který vede k interoperabilitě KOS.

3.2. Aplikace webových technologií

Stejně jako každý jiný obchod a průmysl, také reprezentace a organizace znalostí v síťové éře prošla mnoha dramatickými změnami. Ty zahrnují ne jen metody a technologie používané k tvorbě záznamů dokumentů, ale také standardy, které jsou hlavním klíčem k úspěchu a efektivitě těchto nástrojů v procesu vyhledávání a získávání informací [QIN, 2000]. Tvorba nových a rozvoj stávajících standardů a sémantických webových technologií, konkrétně jejich aplikace na KOS, je jedním z procesů webifikace. Jak již bylo řečeno v kapitole 2.3, webifikace představuje převod existujících KOS do sémantických webových standardů jako jsou na příklad Web Ontology Language (OWL) nebo Simple Knowledge Organization System (SKOS).

3.2.1. SKOS

SKOS³¹ je původním výstupem projektu *Language Independent Metadata Browsing of European Resources* (LIMBER), který byl financován Evropským hospodářským společenstvím³² a částečně financován z programu *Technologie informační společnosti* (Information Society Technologies). Dále byl SKOS vyvíjen v rámci projektu *Semantic Web Advanced Development for Europe* (SWAD-Europe). Poprvé byl SKOS představen na konci roku 2003. Jeho hlavní editoři jsou Alistair Miles a Dan Brickley. Kandidátem na doporučení W3C se SKOS stal na konci roku 2007, návrhem na doporučení v roce 2008. Oficiálně pak bylo doporučení W3C SKOS publikováno v srpnu 2009. Podle Milese [2008] se SKOS zrodil z organizace znalostí a webu.

Slovní spojení *Simple Knowledge Organization System* bylo jako název vybráno proto, aby poukázalo na jeden ze základních cílů tohoto formátu, který spočívá v poskytnutí jednoduchého, přesto účinného rámce pro specifikaci znalostních struktur ve strojově srozumitelné podobě. Rozsah jeho užití se týká ne jen tezaurů, ale i jiných KOS. Cílem je také jeho užití v rámci sémantického webu, který není založen na pouhé výměně dat, ale je i o jejich organizaci distribuovaným a decentralizovaným způsobem [MILES,

³¹ <http://www.w3.org/2004/02/skos/>

³² Evropské hospodářské společenství byla organizace, která existovala v letech 1958-1993.

2005]. SKOS poskytuje jednoduchý, strojově čitelný reprezentační rámec pro KOS. Je flexibilní a rozšiřitelný, aby mohl překlenout rozdíly jednotlivých KOS. SKOS je založen na standardech konsorcia W3C RDF (Resource Description Framework) a RDFS (RDF Schema) a vzhledem k tomu je možné jej používat v kombinaci s dalšími RDF formáty. SKOS byl navržen jako modulární a rozšiřitelný, způsobem, který umožňuje jeho co možná nejjednodušší implementaci a užití.

Základní jednotkou SKOS je koncept, přesněji `<skos:Concept>`, který slouží k reprezentaci pojmů KOS. SKOS nebyl vytvořen k popisu věcí, které nejsou pojmy. To znamená, že může popsat nebo definovat "auto", ale naopak nemůže popsat auto, které představuje skutečný objekt. Každý pojem je jednoznačně identifikován pomocí URI (Uniform Resource Identifier), což umožňuje se na něj odkazovat. URI může vypadat jako obvyklá webová adresa, např. <http://www.example.com/concept>. V nejlepším případě se používají URI tzv. "web-enabled", což znamená, že skutečně odkazují na reprezentaci pojmu přístupnou na webu. Pokud uživatel zadá takovou URI do svého prohlížeče, zobrazí se mu popis pojmu v čitelné podobě, kdežto stroji je prezentován zápis pojmu ve formátu SKOS. Tato vlastnost je velmi důležitým prvkem, který dovoluje propojování dat, což je základním principem sémantického webu.

„SKOS je záměrně navržen tak, aby umožňoval vyjadřovat pouze ty vztahy mezi pojmy, které jsou ve znalostních systémech obvyklé (např. ekvivalence, hierarchie, varianta nebo asociace)“ [MYNARZ, 2009]. Jednotlivé pojmy tak mohou být propojeny s dalšími pojmy v rámci stejného pojmového schématu (v rámci jednoho KOS) pomocí hierarchických linků (`<skos:broader>`, `<skos:narrower>`, `<skos:broaderTransitive>` a `<skos:narrowerTransitive>`) a asociativních linků (`<sko:related>`). Pojmy mohou být opatřeny také poznámkami různých typů (na příklad `<skos:note>`, `<skos:changeNote>`, `<skos:definition>`, `<skos:editorialNote>`, `<skos:example>`, `<skos:historyNote>` nebo `<skos:scopeNote>`).

Užitečným prvkem datového modelu SKOS je možnost vyjádřit shodu nebo podobnost mezi pojmy různých pojmových schémat (v rámci různých KOS). Tyto vazby mohou být vyjádřeny hierarchicky (`<skos:broadMatch>`, `<skos:narrowMatch>`) nebo asociativně (`<skos:relatedMatch>`, `<skos:closeMatch>`, `<skos:exactMatch>`) [PHIPPS, 2010]. Na příklad pomocí elementu `<skos:exactMatch>` lze odkázat na ekvivalentní pojem v jiném pojmovém schématu (jiném KOS), na který lze jednoznačně odkazovat s

použitím URI. Vazby na další pojmy mohou být vytvářeny na všechny znalostní systémy, které jsou k dispozici ve formě prolinkovaných dat.

4. TRADIČNÍ SYSTÉMY ORGANIZACE ZNALOSTÍ NA PŘÍKLADU PSH a LCSH

V kapitole 1.2 byl představen a popsán návrh způsobu hodnocení KOS podle autorů Souzy, Tudhopea a Almeidy [2010]. V této části práce slouží navrhovaný model jako inspirace k popisu dvou tradičních systémů organizace znalostí – Polytematického strukturovaného hesláře (PSH) a Předmětových hesel Kongresové knihovny (LCSH).

4.1.PSH

PSH vznikl ve Státní technické knihovně (dále jen STK)³³ v letech 1991-1993, tedy v letech, kdy se STK stala řešitelem projektu I 096 *Technologie poloautomatické indexace dokumentů s použitím selekčních jazyků verbálního typu pro polytematické fondy*, který byl finančně podpořen programem *Informatizace*. Vznik hesláře souvisel s rychlým rozvojem automatizovaných knihovnických systémů a jejich využíváním v českých knihovnách na počátku 90. let. Další vývoj PSH byl rovněž podpořen ze státního rozpočtu České republiky. Jednalo se prostředky získané z programů výzkumu, rozvoje techniky a technologií schválených usnesením vlády České republiky č. 124 ze dne 17. 4. 1991³⁴, dále finance získané na projekt RS95IF074 *Zajištění dostupnosti informací v knihovnách z hlediska věcného přístupu prostřednictvím polytematického strukturovaného slovníku (tezauru)*, které obdržela STK v rámci programu *INFRA (Budování infrastruktury pro výzkum a vývoj na vysokých školách)*³⁵ a grantové prostředky Ministerstva kultury ČR (dále jen MK) na *Projekt vytvoření anglické verze Polytematického strukturovaného hesláře*³⁶. Od roku 1995 je PSH používán k věcnému popisu v Národní technické knihovně (dále jen NTK) a od počátku roku 1997 je distribuován všem externím zájemcům, jak knihovnám, tak i komerčním firmám. Na podzim roku 2006 byl heslář implementován do autoritní databáze systému Aleph.

³³ Původní název *Státní technická knihovna* se již nepoužívá. Knihovna byla přejmenována a v současné době užívá jméno *Národní technická knihovna*.

³⁴ Finanční prostředky obdržela STK na podporu projektu na základě smlouvy č. I 096/21-3 uzavřené mezi Ministerstvem hospodářství České republiky a STK dne 15. 6. 1993.

³⁵ Jednalo se o smlouvu č. RS95IF074 uzavřenou mezi Ministerstvem školství, mládeže a tělovýchovy a STK dne 7. 8. 1995.

³⁶ STK získala grant MK na překlad PSH do angličtiny v roce 2000.

V současnosti je PSH spravován referátem PSH v NTK a jeho údržba probíhá v prostředí softwaru Aleph 500 ver. 20. Poslední aktualizovaná verze PSH 3.0 z roku 2011 obsahuje přes 13 800 hesel.

4.1.1. Charakteristika PSH podle navrženého modelu – vnitřní vlastnosti

Vnitřní vlastnosti, jak již bylo uvedeno, souvisejí s KOS samotnými a jsou tak chápány jako entity izolované od uživatelů a prostředí. Vnitřní vlastnosti se dále rozdělují na základní a náhodné. Základní vlastnosti zahrnují standardizaci a strukturu dále větvenou na entity a vztahy. Náhodné vlastnosti jsou takové, které úzce souvisí s daným systémem. K náhodným vlastnostem je řazen jazyk, doména, médium a zobrazení. Jazyk a doména jsou v navrženém modelu ještě dále specifikovány.

4.1.1.1. *Struktura, entity a vztahy PSH*

PSH tvoří stromovou strukturu obsahující 44 základních hesel. Každé heslo má k sobě přiřazenou dvouznakovou zkratku, jež označuje příslušnost hesla do dané tematické řady³⁷. Díky tomu lze u hesel stojících na nižším stupni hierarchie zjistit příslušnost k dané tematické řadě.

Základní lexikální jednotkou PSH je předmětové heslo. Nejjednodušší formu předmětového hesla představuje jednoslovné substantivum. Předmětová hesla mohou také tvořit víceslovná spojení, na příklad spojení adjektiv a substantiv (např. *astronomické přístroje*), zřídka se v hesláři vyskytují hesla tvořené frází se spojkou (např. *elektrina a magnetismus*). Zpočátku byla u víceslovných lexikálních jednotek preferována substantivní inverze. V současné době platí pro víceslovná spojení pravidlo dodržování přirozeného pořádku slov. Do PSH nejsou zařazována hesla geografická, personální, chronologická, hesla označující profesní skupiny nebo profesní organizace a názvy produktů (např. Škoda).

V PSH se vyskytují tři základní typy vztahů. Jedná se o hierarchické, asociační a ekvivalenční vztahy. Pro hierarchickou strukturu je charakteristický vztah nadřazenosti

³⁷ Např. heslo "matematika" *ma*, heslo "oxidy" *ch*, heslo "savci" *bi*.

a podřazenosti³⁸. Vztahy asociace se v hesláři vyskytují v případech, kdy je vhodné upozornit uživatele na další příbuzná (asociovaná) hesla, která jsou umístěna v jiné části hesláře. Jsou řešena přidruženým odkazem "viz též"³⁹. Ekvivalenční vztahy představují vazbu mezi heslem a nepreferovaným termínem. Řeší ji vylučovací odkaz "viz"⁴⁰.

4.1.1.2. Standardizace PSH

PSH je vytvářen, spravován a rozvíjen dle *Pravidel pro správu a aktualizaci Polytematického strukturovaného hesláře (PSH)* a dokumentu *Zásady pro vytváření Polytematického strukturovaného hesláře (PSH)*. Při indexaci dokumentů jsou zároveň respektována *Indexační pravidla pro práci s Polytematickým strukturovaným heslářem (PSH)*.

4.1.1.3. Jazyk PSH

PSH je česko-anglický kontrolovaný slovník, který je vytvářen v souladu s principy přirozeného jazyka. Nejprve byl vytvořen jako český slovník hesel, později, v roce 2000 byl na základě grantu přeložen do angličtiny. Tento krok z něj udělal systém zařaditelný mezi mezinárodně srozumitelné a použitelné nástroje i v cizojazyčných knihovnách.

PSH je dostupný ve dvou formátech. Základním formátem PSH, který je určený především pro knihovny, je formát MARC 21 pro authority⁴¹, který se používá v automatizovaných knihovnických systémech, ale pro webovou distribuci se nehodí. Příklad zobrazení záznamu hesla PSH ve formátu MARC 21 pro authority je k dispozici na obr. č. 7. Druhým formátem PSH je formát SKOS (popsaný v kapitole 3.2.1). Na obr. č. 8 je zachycen příklad zobrazení téhož hesla "jaderná fyzika" tentokrát ve formátu SKOS.

³⁸ Např. heslo "algebra" je podřazeno heslu "matematika" a naopak platí, že heslo "matematika" je nadřazeno heslu "algebra".

³⁹ Např. heslo "rovnice" viz též "diferenciální rovnice".

⁴⁰ Např. heslo "rovnice" viz "nelineární rovnice".

⁴¹ <http://www.loc.gov/marc/authority/>

Name Tag	Data
*****	Záznam dokumentu 1 z 1
FMT	VA
LDR	-----nz--a22-----n--4500
001	PSH3634
003	CZ-PrSTK
005	20070126161437.0
008	070126na ann bab ----- -a a-----
040	a ABA013 b cze
150	a jaderná fyzika x fy
550	a jaderné reaktory x en
550	a jaderná chemie x ch
5501	w h a rozptyl částic x fy
5501	w h a svazky nabitých částic x fy
5501	w h a urychlovače nabitých částic x fy
5501	w h a atomové jádro x fy
5501	w h a fyzika nízkých energií x fy
5501	w h a fyzika vysokých energií x fy
5501	w h a elementární částice x fy
5509	w g a fyzika x fy
75007	a nuclear physics 2 epsh
CAT	c 20070127 l STK10 h 0945
SYS	000003635

Obr. 7: Příklad zobrazení záznamu hesla "jaderná fyzika" ve formátu MARC 21 pro autority

```

<rdf:RDF>
- <skos:Concept rdf:about="http://psh.ntkcz.cz/skos/PSH3634">
  <skos:prefLabel xml:lang="cs">jaderná fyzika</skos:prefLabel>
  <skos:prefLabel xml:lang="en">nuclear physics</skos:prefLabel>
  <skos:related rdf:resource="http://psh.ntkcz.cz/skos/PSH2543"/>
  <skos:related rdf:resource="http://psh.ntkcz.cz/skos/PSH5645"/>
  <skos:broader rdf:resource="http://psh.ntkcz.cz/skos/PSH2910"/>
  <skos:narrower rdf:resource="http://psh.ntkcz.cz/skos/PSH3639"/>
  <skos:narrower rdf:resource="http://psh.ntkcz.cz/skos/PSH3694"/>
  <skos:narrower rdf:resource="http://psh.ntkcz.cz/skos/PSH3695"/>
  <skos:narrower rdf:resource="http://psh.ntkcz.cz/skos/PSH3693"/>
  <skos:narrower rdf:resource="http://psh.ntkcz.cz/skos/PSH3649"/>
  <skos:narrower rdf:resource="http://psh.ntkcz.cz/skos/PSH3635"/>
  <skos:narrower rdf:resource="http://psh.ntkcz.cz/skos/PSH3636"/>
  <skos:inScheme rdf:resource="http://psh.ntkcz.cz/skos"/>
</skos:Concept>
</rdf:RDF>

```

Obr. 8: Příklad zobrazení záznamu hesla "jaderná fyzika" ve formátu SKOS

4.1.1.4. Doména PSH

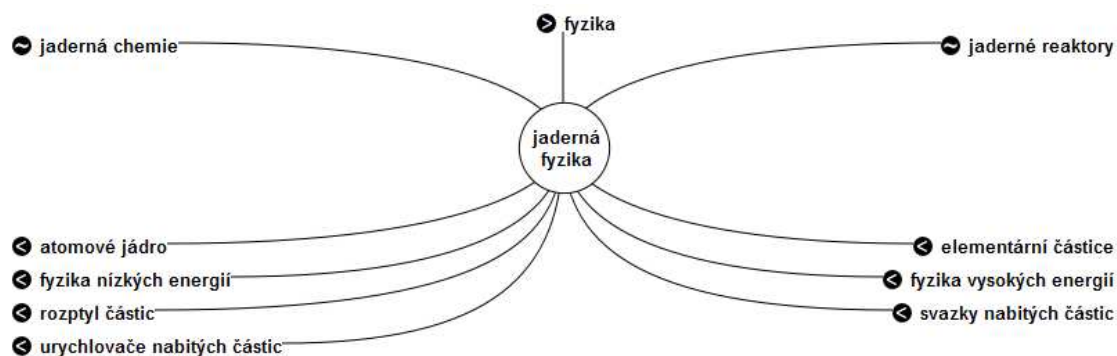
Jak napovídá celý název PSH, jedná se o polytematický systém, obsahující 44 základních hesel ze všech základních oblastí lidského poznání. Každé heslo je v rámci oboru podle svého obsahu a míry obecnosti zařazeno do šesti-, výjimečně sedmistupňové hierarchie (1-7 úrovní). Hesla příliš specifická a hesla patřící do hlubších vrstev logické struktury nemají v PSH místo. Při tvorbě nových hesel je nutné dodržovat stejnou úroveň specifičnosti.

4.1.1.5. Média a zobrazení PSH

PSH jako celek nebyl nikdy publikován v tištěné podobě, k dispozici byl vždy v podobě elektronické. V současné době se stránkách Národní technické knihovny v sekci PSH se pod odkazem "PROHLÍŽENÍ HESLÁŘE"⁴² nachází webové rozhraní pro grafické prohlížení PSH. Na hlavní stránce je vyobrazen seznam 44 hlavních tematických řad PSH. Systém hypertextových odkazů umožňuje procházet heslářem po hierarchické a asociativní ose. Hlavní navigaci ve stromové struktuře představují preferovaná znění hesel, k dispozici jsou ale i hesla nepreferovaná nebo jejich anglické ekvivalenty.

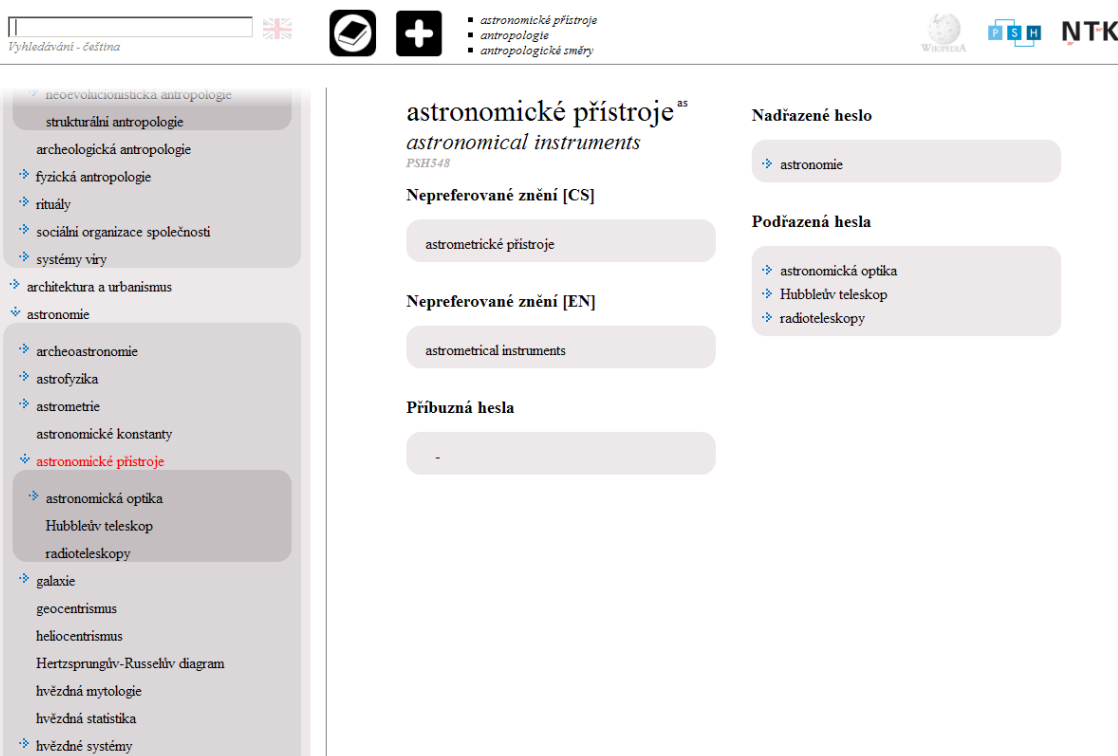
⁴² <http://psh.ntkcz.cz/skos/>

Vizualizace je ztvárněna středovým kruhem, ve kterém se nachází hledaný termín s připojenými vazbami na obvodu. Vztahy okolních hesel jsou indikovány malými ikonami symbolizujícími nadřazenost, podřazenost nebo podobnost úrovně PSH (viz obr. č. 9). Pod zobrazením jednotlivých hesel se nachází seznam dokumentů z katalogu NTK, které byly při katalogizaci daným termínem indexovány.



Obr. 9: Příklad grafického zobrazení hesla "jaderná fyzika" na stránkách "PROHLÍŽENÍ HESLÁŘE"

Dalším příkladem zobrazení PSH je webové rozhraní *PSH Manager Online* (viz obr. č. 10), které bylo vytvořeno dle návrhů pracovníků věcné katalogizace v NTK za účelem zefektivnění práce s heslářem při indexaci dokumentů. Tato aplikace je v současné době přístupná pouze v budově NTK, v budoucnu se však počítá s jejím zpřístupněním všem spolupracujícím knihovnám. Výhodou *PSH Manager Online* je možnost užití funkcí našeptávače (český i anglický), zobrazení historie prohlížení (3 poslední hesla) a funkce automatického odkazování na odpovídající hesla v portálu české Wikipedie.



Obr. 10: Příklad zobrazení hesla "astronomické přístroje" v rozhraní PSH Manager Online

4.1.2. Charakteristika PSH podle navrženého modelu – vnější vlastnosti

Vnější parametry hodnocení KOS souvisejí s prostředím, ve kterém jsou tyto systémy užívány. Zahrnují účel, komunitu uživatelů, závislost a aktualizace.

4.1.2.1. Účel PSH

Účelem PSH je primárně indexace dokumentů. „Jedním ze záměrů tvůrců PSH bylo nahradit univerzální klasifikační systémy typu MDT podstatně otevřenějším a uživatelsky přívětivějším selekčním jazykem, blízkým přirozenému jazyku. Ten měl jednak usnadnit indexaci především univerzálních a polytematických knihovních fondů, jednak umožnit uživatelům efektivní prohledávání v automatizovaných katalozích typu OPAC“ [SMOLKA, 1999]. V době svého vzniku byla účelem PSH především indexace dokumentů v tehdejší STK a dalších spolupracujících českých knihovnách. V posledních třech letech byla využitelnost PSH posunuta za hranice knihoven. Současným záměrem NTK je rozšíření PSH i mezi komunitu běžných uživatelů

internetu, kteří jej mohou využít na příklad pro předmětovou indexaci webových dokumentů.

4.1.2.2. Komunita uživatelů PSH

Komunita uživatelů je úzce provázána s účelem celého systému. V případě PSH, jsou uživateli především knihovny. V současné době patří mezi aktivní uživatele PSH kromě Národní technické knihovny také na příklad knihovny Českého vysokého učení technického v Praze, Ústřední knihovna Vysokého učení technického v Brně, Vědecká knihovna v Olomouci, knihovna Západočeského muzea v Plzni či Ústřední knihovna Filozoficko-přírodovědecké fakulty Slezské univerzity v Opavě. V roce 2010 byl PSH nově implementován do systému Knihovny Bankovního institutu vysoká škola (BIVŠ, a.s.). Jednotlivci se mohou stát uživateli PSH zprostředkovaně, např. při vyhledávání v katalogu knihovny, která PSH využívá. Novým uživatelem PSH se také stala skupina IT odborníků z Pardubické univerzity, která vytváří testovací rozhraní pro vyhledávání zaměstnanců univerzity dle jejich zaměření podle hierarchie hesel PSH.

4.1.2.3. Aktualizace PSH

Rychlé zastarávání termínů je typickým problémem všech heslářů, které mají pevnou strukturu i obsah. Tvorba hesláře se tak stává nikdy nekončící činností. U správy tezauru dříve nebo později dospějeme do bodu, kdy je nezbytné přidat nové termíny a naopak zastaralé termíny smazat. Stejně tak je potřeba doplnit instrukce a pravidla pro správu takového tezauru. „*Všechny změny však musí být opatrně zvládnuty, pokud nechceme, aby se tezaurus proměnil v chaos*“ [BROUGHTON, 2006, s. 197]. V otázce časové frekvence aktualizace se střetávají dva protichůdné požadavky; na jedné straně by řízený slovník měl být stabilní, na straně druhé by ale měl být aktualizován dle potřeby.

Klíčovým kritériem aktualizace PSH je zachování kontinuity. V hesláři by neměly být prováděny žádné výrazné změny, preferovány jsou pouze změny dílčího charakteru. Nedoporučuje se proto zbytečně měnit hierarchickou strukturu a vztahy mezi hesly a přesouvat pojmy mezi tematickými řadami. Stabilita je v tomto případě důležitější než snaha o vylepšení detailů. Ne všechna hesla mohou být do nové verze přijata. Zejména je třeba dbát na skutečnost, že hesla příliš specifická a hesla patřící do hlubších vrstev logické struktury nemají v PSH místo. V případě nezbytnosti lze tato hesla zařadit

do skupiny odkazů. V zásadě tedy platí, že při zařazování nových hesel do struktury by měla být dodržena úroveň obecnosti. Při aktualizaci PSH je téměř vyloučeno vyřazovat deskriptory použité ve starších verzích. V hesláři je možné provádět pouze tyto typy úprav – zařazování nových hesel (deskriptorů), která jsou výsledkem pokroku lidského poznání a tvorba vylučovacích a přidružovacích odkazů. Vylučovací odkazy se používají pro možné vyloučení synonym nebo významově příliš podřadných hesel. Přidružovací odkazy vyjadřují asociativní vztahy mezi pojmy tak, aby pomocí odkazů umožnily nalézt obsahově příbuzné dokumenty.

V minulosti byly aktualizace PSH prováděny nahodile a nepravidelně. Vývoj aktualizací lze spatřovat ve vzniku číslovaných verzí PSH (např. PSH 1.4, PSH 2.0). Tyto aktualizace probíhaly především ve spolupráci s odborníky s různých oborů. Postup spočíval ve výběru odborníka, který pak navrhoval úpravy v jednotlivých částech PSH a poté byly jeho návrhy schvalovány pracovníky knihovny. Nyní je proces aktualizace PSH dle *Pravidel pro správu a aktualizaci PSH* optimalizován třemi způsoby.

Prvním z nich je export a následná analýza pole pro volně tvořený termín, které se v NTK používá při věcném popisu knih (obsah pole 653 v bibliografickém záznamu ve formátu MARC 21). Pole 653 je v NTK používáno doplňkově, a to v případech, kdy danou skutečnost nelze dostatečně přesně vyjádřit prostřednictvím hesel PSH (jedná se o podrobnější úroveň zpracování tématu) nebo se v hesláři odpovídající heslo nevyskytuje (jedná se o nový termín).

Druhým způsobem optimalizace je analýza logu vyhledávání. Ta spočívá v rozboru uživatelských dotazů zadávaných při vyhledávání v katalogu NTK. „*Log soubory, zkráceně logy, jsou textové soubory obsahující záznamy o činnosti nějaké konkrétní aplikace. V případě webových serverů jsou do logů ukládány veškeré požadavky, které byly na server vzneseny. Zpětnou analýzou těchto dat pak můžeme zjišťovat cenné informace o fungování sledovaného webu.*“ [SMRT, 2007]. Log vyhledávání lze charakterizovat také jako soubor, do kterého se zaznamenávají uživatelské požadavky společně s textovým záznamem komunikace mezi zúčastněnými počítačovými vrstvami a systémovými informacemi. Analýza webových logů je nejčastěji spojována se statistikou počtu návštěvníků, průměrné doby jejich setrvání, statistiky přístupů apod. Cílem analýzy logu v referátu PSH v NTK je optimalizace aktualizace PSH, která by

měla usnadnit uživatelům vyhledávání a přiblížit se jejich potřebám. Rozbor dotazů pokládaných uživateli při práci v online katalogu může do určité míry vyřešit disproporci mezi termíny používanými při indexaci a termíny používanými samotnými uživateli popisujícími jejich informační potřeby. Záznam logu může mít až milion řádků, proto je v první fázi analýzy nutné odfiltrovat nepotřebný text. Tento proces využívá schopností regulárních výrazů, které představují univerzální, uživatelsky nadefinovanou šablonu struktury textu. Při nalezení části textu shodné s regulárním výrazem je možné tato data extrahovat a výsledky použít pro jednoduchou statistickou analýzu. Cílem algoritmu jsou termíny zadávané při vyhledávání ve všech polích, v poli předmětovém a v heslech PSH. Získané termíny se periodicky přidávají do výstupního souboru. Při vyšší frekvenci některého z nich je heslo pracovníky referátu PSH posuzováno jako kandidát na nové heslo PSH. K výhodám analýzy logu patří možnost jejího jednorázového provedení za delší období a fakt, že větší objem dat poskytne statisticky významnější výsledky. Sledování uživatelského chování je velmi užitečné, jelikož názorně ukazuje, jakým způsobem se uživatelé v katalogu NTK pohybují, jaký postup uplatňují při hledání určitého dokumentu a jak se jim to daří.

Třetím, doplňkovým způsobem získávání nových hesel PSH je webový formulář pro zasílání návrhů nového hesla PSH (viz obr. č. 11), který byl vytvořen v souvislosti s přípravou aktualizace PSH a řešením otázky, jak mohou uživatelské knihovny a odborná veřejnost přispět k revizím PSH. Formulář mohou využívat jak katalogizátoři oddělení věcného zpracování v NTK, tak externí uživatelé, např. zájemci o problematiku selekčních jazyků, ale i zákazníci NTK. Po vyplnění příslušných polí je tabulka odeslána pracovníkům referátu PSH, kteří získané návrhy posuzují, kontrolují případnou duplicitu a zvažují jejich zařazení do struktury PSH.

Návrh nového hesla PSH

Nové heslo PSH

Synonymum nového hesla
(česky)

Nové anglické heslo PSH

Synonymum nového hesla
(anglicky)

E-mail

ISBN nebo ISSN

Poznámka

Potvrdit odeslání

Obr. 11: Webový formulář pro zasílání návrhů nového hesla PSH

4.2.LCSH

Předmětová hesla Kongresové knihovny (dále jen LCSH) byla poprvé publikována již v roce 1914 pod názvem *Subject Headings Used in the Dictionary Catalogues of the Library of Congress*⁴³. Při jejich vzniku bylo čerpáno z díla *Pravidla pro slovníkový katalog (Rules for a Dictionary Catalog)*⁴⁴ autora Charlese Ammiho Cuttera. V roce 2011 bylo publikováno již 33. šestisvazkové tištěné vydání LCSH zahrnující více než 337 000 předmětových hesel a odkazů [LIBRARY OF CONGRESS, 2011a]. V současné době patří LCSH k nejpoužívanějším předmětovým heslům na světě [CHOWDHURY G. G.; CHOWDHURY S., 2007].

⁴³ Pod tímto názvem vycházela Předmětová hesla Kongresové knihovny až do roku 1975, kdy byl titul změněn na *Library of Congress Subject Headings*.

⁴⁴ Dílo bylo publikováno v roce 1904.

4.2.1. Charakteristika LCSH podle navrženého modelu – vnitřní vlastnosti

Vnitřní vlastnosti souvisejí s KOS samotnými a jsou tak chápány jako entity izolované od uživatelů a prostředí. Vnitřní vlastnosti se dále rozdělují na základní a náhodné. Základní vlastnosti zahrnují standardizaci a strukturu dále větvenou na entity a vztahy. Náhodné vlastnosti jsou takové, které úzce souvisí s daným systémem. K náhodným vlastnostem je řazen jazyk, doména, médium a zobrazení. Jazyk a doména jsou v navrženém modelu ještě dále specifikovány.

4.2.1.1. *Struktura, entity a vztahy LCSH*

LCSH je kontrolovaný slovník prekoordinovaného typu. Základní lexikální jednotkou tohoto systému je předmětové heslo. Nejjednodušší formou předmětového hesla je jednoslovné substantivum. Předmětová hesla však mohou být vyjádřena také spojením adjektiva a substantiva (např. *Chemical engineering*), předložkovou vazbou (např. *Practice of law*) nebo může být heslo tvořeno frází se spojkou (např. *Law and socialism*). Zpočátku byla u víceslovných lexikálních jednotek preferována substantivní inverze. V současném systému LCSH se inverzní tvary nepodporují, slouží pouze jako variantní vstup pro koncového uživatele (odkaz typu viz) [BALÍKOVÁ, 2001].

Součástí hesel mohou být kvalifikátory, které slouží k tomu, aby se zabránilo problémům spojeným s homonymií. Tyto kvalifikátory se často využívají také k vysvětlení ne příliš známého termínu. Některé z nich mohou jednoduše odkazovat na disciplínu, do které heslo náleží (např. *Outgassing (Low pressure environments)*).

Podle Balíkové [2001] bývá v odborné literatuře LCSH vytýkán malý výskyt specifických termínů. Potřebné specifičnosti pro vyjádření detailnějšího tématu dokumentu bývá dosaženo použitím řetězce předmětového hesla, který je specifikem LCSH. Jedná se o víceprvková předmětová hesla, která se při vyjádření komplexního tématu skládají z několika prvků (např. *Natural gas—Law and legislation*). Informace uvedená jako první v pořadí je rozvíjena dalšími prvky předmětového hesla, které mohou vyjadřovat tematické, geografické, chronologické a formální okolnosti. Užití řetězců předmětových hesel ale není vhodné na příklad při aplikaci LCSH v plnotextových databázích. Obecnost předmětových hesel LCSH lze do určité míry

vyřešit odstraněním těchto řetězců a jejich nahrazením specifitějším heslem. Na příklad původní heslo *Gas - Pipe lines* bylo nahrazeno *Gas pipelines*.

V LCSH se vyskytují, stejně jako v PSH, vztahy hierarchie, asociace a ekvivalence. K zachycení těchto vztahů jsou užívány dvouznakové zkratky UF (Used for) pro ekvivalenční vztahy⁴⁵, BT (Broader topic) a NT (Narrower topic) pro vyjádření hierarchických vztahů⁴⁶ a RT (Related topic) pro asociační vztahy⁴⁷. Kromě asociačních odkazů na příbuzné termíny se v LCSH vyskytují také obecné odkazy, které neodkazují na jedno specifické předmětové heslo, ale odkazují na celou skupinu hesel.

4.2.1.2. Standardizace LCSH

Základním dokumentem k užití LCSH je manuál *Subject Cataloging Manual: Subject Headings*, který zahrnuje stejné instrukce, které jsou užívány katalogizátory v Kongresové knihovně. Přestože se v manuálu mohou vyskytnout některé instrukce k procesům vlastním pouze Kongresové knihovně, většina pravidel je základních a aplikovatelných pro všechny, kteří chtějí LCSH užívat.

4.2.1.3. Jazyk LCSH

LCSH je jednojazyčný. Obsahuje hesla pouze v angličtině⁴⁸. LCSH je dostupný ve dvou formátech. Základním formátem LCSH je MARC 21 pro autority⁴⁹. Příklad zobrazení záznamu hesla "nukleární energie" ve formátu MARC 21 pro autority je k dispozici na obr. č. 12. Druhým formátem LCSH je formát SKOS (popsaný v kapitole 3.2.1). Na obr. č. 13 je zachycen příklad zobrazení hesla "jaderný průmysl" tentokrát ve formátu SKOS.

⁴⁵ Např. heslo "nuclear energy" UF (viz) "atomic energy".

⁴⁶ Např. heslo "nuclear industry" je podřazeno heslu "nuclear energy" a naopak platí, že heslo "nuclear energy" je nadřazeno heslu "nuclear industry".

⁴⁷ Např. heslo "nuclear energy" RT (viz též) "nuclear power plants".

⁴⁸ Existují však projekty, které LCSH překládají do jiných jazyků na příklad do španělštiny (viz <http://lcsch-es.org/index.html>). Tento projekt vznikl proto, aby byl podpořen předmětový přístup k jednotnému online katalogu deseti univerzit v Mexico City, a zároveň tento nástroj může sloužit bilingvním komunitám ve Spojených státech amerických [STONE, 2000].

⁴⁹ <http://www.loc.gov/marc/authority/>

Subject Record [Nuclear energy]

ID:	sh 85092958	Entered:	041006	Replaced:	20041104
008/06 Geo Subd:	i-Indirect	008/11 SH System:	a-LCSH	008/29 Ref Eval:	b-Not eval
008/07 Roman:	-No attempt	008/15 Subj Use:	a-Appropriate	008/31 Rec Upd:	a-Can be used
008/09 Kind Rec:	a-Estab hdg	008/17 Type Subd:	n-Not applic	008/33 Level Estab:	a-Fully

010 \$a sh 85092958
 040 \$a DLC \$c DLC \$d DLC
 053 0 \$a QC791.9 \$b QC792.8 \$c Physics
 053 0 \$a TK9001 \$b TK9401 \$c Technology
 150 \$a Nuclear energy
 450 \$w nne \$a Atomic energy
 450 \$w nne \$a Atomic power
 450 \$a Energy, Atomic
 450 \$a Energy, Nuclear
 450 \$a Nuclear power
 450 \$a Power, Atomic
 450 \$a Power, Nuclear
 550 \$w g \$a Force and energy
 550 \$w g \$a Nuclear physics
 550 \$w g \$a Power resources
 550 \$a Nuclear engineering
 550 \$a Nuclear facilities
 550 \$a Nuclear power plants

Obr. 12: Příklad zobrazení záznamu hesla "nukleární energie" (Nuclear energy) ve formátu MARC 21 pro autority v rozhraní klasifikačního webu

```

- <rdf:RDF>
- <rdf:Description rdf:about="http://id.loc.gov/authorities/sh85092998#concept">
  <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
  <skos:broader rdf:resource="http://id.loc.gov/authorities/sh85092958#concept"/>
  <skos:broader rdf:resource="http://id.loc.gov/authorities/sh85043141#concept"/>
  <owl:sameAs rdf:resource="info:lc/authorities/sh85092998"/>
  <skos:closeMatch rdf:resource="http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb11932104m"/>
  <dcterms:created rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">1986-02-11T00:00:00-04:00</dcterms:created>
  <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">1989-03-31T13:42:59-04:00</dcterms:modified>
  <skos:inScheme rdf:resource="http://id.loc.gov/authorities#topicalTerms"/>
  <skos:inScheme rdf:resource="http://id.loc.gov/authorities#conceptScheme"/>
  <skos:prefLabel xml:lang="en">Nuclear industry</skos:prefLabel>
  <skos:altLabel xml:lang="en">Atomic energy industries</skos:altLabel>
  <skos:altLabel xml:lang="en">Nuclear energy industry</skos:altLabel>
  <skos:altLabel xml:lang="en">Atomic power industry</skos:altLabel>
  <skos:altLabel xml:lang="en">Atomic industry</skos:altLabel>
  <skos:altLabel xml:lang="en">Nuclear power industry</skos:altLabel>
</rdf:Description>
- <rdf:Description rdf:about="http://id.loc.gov/authorities/sh85092958#concept">
  <skos:prefLabel xml:lang="en">Nuclear energy</skos:prefLabel>
</rdf:Description>
- <rdf:Description rdf:about="http://id.loc.gov/authorities/sh85043141#concept">
  <skos:prefLabel xml:lang="en">Energy industries</skos:prefLabel>
</rdf:Description>
</rdf:RDF>
  
```


4.2.1.4. Doména LCSH

LCSH je všeobecný systém, nicméně vychází také z principu založeném na vlastnosti "literary warrant" [CHOWDHURY G. G.; CHOWDHURY S., 2007]. Znamená to, že LCSH reflektuje skutečná témata dokumentů vyskytující se ve fondu Kongresové knihovny a zároveň odráží terminologii používanou v současné literatuře. Literary warrant představuje v kontextu organizace znalostí koncept, který klade důraz na to, aby nová předmětová hesla reprezentovala témata, která se vyskytují v konkrétním fondu.

4.2.1.5. Média a zobrazení LCSH

LCSH je dostupný v tištěné podobě, v několika svazcích, které jsou v literatuře označovány jako tzv. "červené knihy"⁵⁰. Kromě klasických tištěných svazků je LCSH dostupný také v elektronické podobě.

LCSH je možné zobrazit na příklad prostřednictvím Klasifikačního webu, který umožňuje prohlížení a vyhledávání předmětových hesel (viz obr. č. 14). Na obr. č. 15 je vidět způsob zobrazení hesla "jaderná energie" v rozhraní klasifikačního webu. Výhodou jeho využití je možnost zobrazení korelace vybraného předmětového hesla s Deweyho desetinným tříděním a Tříděním Kongresové knihovny. Stejně tak je zde k dispozici funkce zobrazení hesla ve formátu MARC 21. Přístup do této databáze je placený a je podmíněn heslem.

Dalším způsobem zobrazení LCSH jsou stránky Kongresové knihovny *Authorities & Vocabularies*⁵¹. Zadáním hledaného termínu do vyhledávacího okna se v případě kladné odezvy systému zobrazí požadované heslo, u něhož je možné zobrazit vizualizaci (viz obr. č. 16).

⁵⁰ Přeloženo z anglického spojení "red books". Toto označení vzniklo pravděpodobně na základě skutečné červené barvy, které je typická pro přebal těchto svazků.

⁵¹ <http://id.loc.gov/authorities>



Now with LC Names

Main Menu

[Browse LC Classification Schedules](#)[LC Classification Search](#)[+ Browse LC Subject Headings](#)[Search LC Subject Headings](#)[Search Genre/Form Headings](#)[Search Children's Subject Headings](#)[+ Browse LC Name Headings](#)[+ Search LC Name Headings](#)[★ Bibliographic Correlations \(menu\)](#)[Log Out](#) - [Preferences](#) - [Outline](#) - [Subsets](#) - [Bookmarks](#) - [Help](#)
[Auto Login Menu](#) - [Diagnostics](#) - [Legal Notices](#) - [Contacts](#) - [User Group](#) - [Home](#)[Standard Classification Browser](#) - [Enhanced Classification Browser](#) - [Hierarchy Classification Browser](#)[+ New feature](#) [★ Major upgrade](#)

- Subclass KBS is under development and not yet approved for use. For further information see the note that appears above KBS3
● Dewey, Dewey Decimal Classification and WebDewey are registered trademarks of OCLC Online Computer Library Center, Inc.

Available from the [Cataloging Distribution Service](#), Library of Congress*Obr. 14: Klasifikační web: hlavní menu*

LC Subject Search: Structured subject heading

Nuclear energy (May Subd Geog) ☐

[\[QC791.9-792.8 \(Physics\)\]](#)

[\[TK9001-9401 \(Technology\)\]](#)

UF Atomic energy [Former Heading]

Atomic power [Former Heading]

Energy, Atomic

Energy, Nuclear

Nuclear power

Power, Atomic

Power, Nuclear

BT [Force and energy](#)

[Nuclear physics](#)

[Power resources](#)

RT [Nuclear engineering](#)

[Nuclear facilities](#)

[Nuclear power plants](#)

NT [Binding energy](#)

[Nuclear batteries](#)

[Nuclear excavation](#)

[Nuclear industry](#)

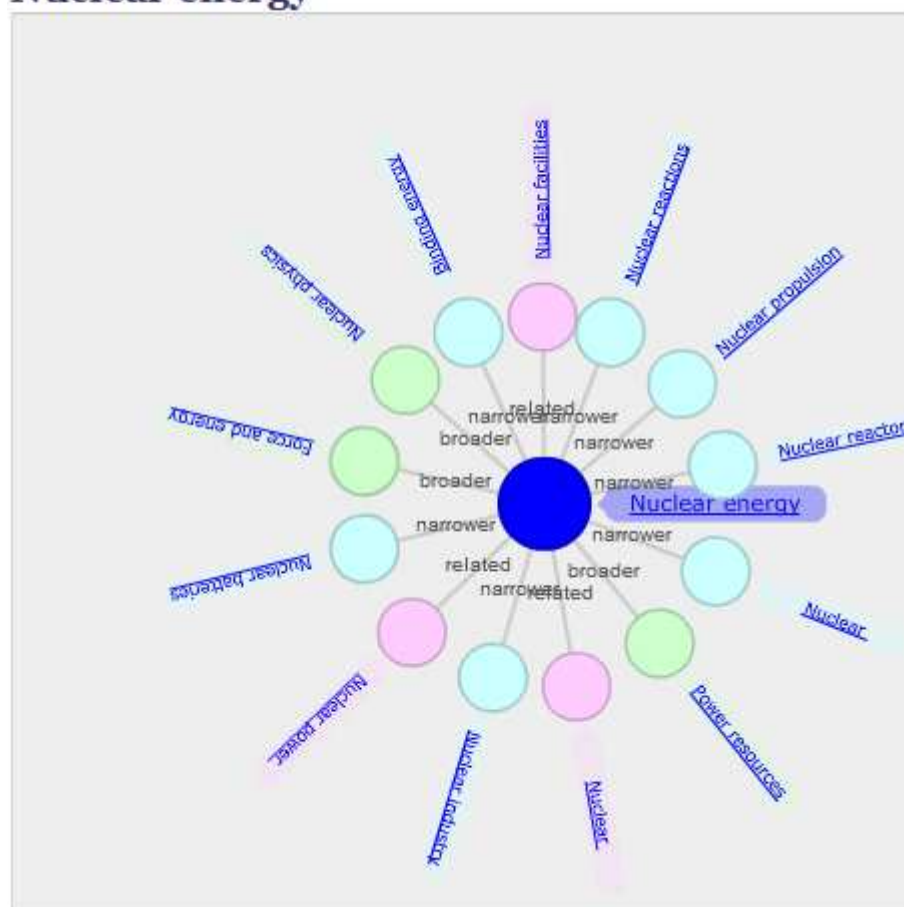
[Nuclear propulsion](#)

[Nuclear reactions](#)

[Nuclear reactors](#)

Obr. 15: Příklad zobrazení záznamu hesla "jaderná energie" (Nuclear energy) v rozhraní Klasifikačního webu

Nuclear energy



Obr. 16: Příklad vizualizace hesla "jaderná energie" (Nuclear energy)

4.2.2. Charakteristika LCSH podle navrženého modelu – vnější vlastnosti

Vnější parametry hodnocení souvisejí s prostředím, ve kterém jsou KOS užívány. Zahrnují účel, komunitu uživatelů, závislost a aktualizace.

4.2.2.1. Účel LCSH

Primárním účelem LCSH je reprezentace předmětu dokumentů, které jsou součástí sbírek Kongresové knihovny Spojených států amerických⁵². Systém je však využíván také dalšími knihovnami a institucemi, které se zabývají indexací. Zpočátku byl využíván především ve velkých veřejných a univerzitních knihovnách a akademických institucích. Později, ve třicátých dvacátého století, se rozšířil do dalších knihoven

⁵² <http://www.loc.gov/index.html>

[STONE, 2000]. V současné době tedy Kongresová knihovna plní funkci autority v tvorbě předmětových hesel a poskytuje knihovnické komunitě řadu příruček, školení a doporučení, jak LCSH aplikovat.

4.2.2.2. Komunita uživatelů LCSH

Uživatelem LCSH je kromě Kongresové knihovny celá řada velkých, malých, veřejných i vědeckých knihoven nejen ve Spojených státech amerických, ale i v dalších zemích. „*Tento pořádací nástroj se rozšířil nejenom ve všech anglicky mluvících zemích, ale postupně se začal používat i v neanglickém prostředí - Španělsku, Portugalsku, Řecku, Francii, Maďarsku, Litvě*“ [BALÍKOVÁ, 2001]. LCSH je v současné době využíván také v neknihovnickém prostředí na příklad v akademické virtuální knihovně *INFOMINE*⁵³, ve které je LCSH využíván k indexaci informačních zdrojů⁵⁴.

4.2.2.3. Aktualizace LCSH

Tezaury, klasifikační schémata, kontrolované slovníky, předmětová hesla, ale i další typy KOS musí být neustále aktuální. Jestliže je daný systém užíván k indexaci dokumentů pouze v jedné databázi nebo v rámci jedné knihovny, proces aktualizace je relativně jednoduchý. Pokud se ale na straně druhé jedná o systém široce využívaný, musí se správci systému ohlížet také na jeho ostatní uživatele.

Návrhy na změny tak mohou podávat nejen katalogizátoři Kongresové knihovny a spolupracující instituce zapojené do kooperativního programu, ale i ostatní uživatelé a zájemci o problematiku LCSH. Navrhovat změny nebo nová předmětová hesla lze na příklad prostřednictvím webového formuláře (viz obr. č. 17). Nově navržená předmětová hesla a nebo návrhy na úpravu existujících hesel jsou hodnoceny, popř. modifikovány a následně schvalovány vybranou skupinou, která se schází jednou týdně. Jestliže členové schůzky dospějí k zásadnímu rozhodnutí, komunikují jej

⁵³ <http://infomine.ucr.edu/>

⁵⁴ INFOMINE je služba, která poskytuje přístup k několika tisícům webových zdrojů zahrnujících databáze, elektronické časopisy, sborníky apod. Služba byla zprovozněna v lednu 1994 jako projekt Knihovny Kalifornské univerzity v Riverside [CHOWDHURY G. G.; CHOWDHURY S., 2007].

prostřednictvím zprávy, která je veřejně publikována jednou týdně [LIBRARY OF CONGRESS, 2011].

Details

Suggest Terminology

Suggest terminology

The LC Authorities and Vocabularies service welcomes any suggestions you might have about terminology used for a given heading or concept.

Would you like to suggest a change to this heading?

- [Library of Congress Subject Headings](#)

Please provide your name, email, and your suggestion so that we can begin assessing any terminology changes.

Alternatively, you can contact us directly at ldfeedback@loc.gov.

Fields denoted with an asterisk (*) are required.

* Name

* E-mail

* Suggested terminology

I recommend changing the terminology for <http://id.loc.gov/authorities#conceptscheme> from "Library of Congress Subject Headings" to:

Send

Cancel

[Privacy policy](#)

Obr. 17: Webový formulář pro zasílání návrhů na změny nebo nová předmětová hesla LCSH

Proces schvalování nových termínů pak může vyžadovat rozšířené a komplexnější konzultace. Změny totiž mohou mít dopad na mnoho knihoven spolu s jejich koncovými uživateli. Indexátoři proto musí být o změnách včas informováni, což v případě světově rozšířeného systému, jako je LCSH, představuje nelehký úkol.

LCSH je aktualizován týdně. Do nové verze LCSH je každoročně přidáno přibližně 6000-8000 nových hesel. Informace o nových, změněných nebo zrušených předmětových heslech lze získat prostřednictvím několika zdrojů. Jednak je vydáván týdenní seznam *Library of Congress Subject Headings (LCSH) Weekly Lists*, čtvrtletně

je potom tento soupis shrnut a vydán v tištěné podobě v rámci *Cataloging Service Bulletin*, který je rovněž dostupný v elektronické podobě⁵⁵.

⁵⁵ <http://www.loc.gov/cds/PDFdownloads/csb/>

5. ANALÝZA VYBRANÝCH TRENDŮ VE VÝVOJI SYSTÉMŮ ORGANIZACE ZNALOSTÍ NA PŘÍKLADECH PSH A LCSH

Tato kapitola je zaměřená na analýzu vybraných trendů ve vývoji KOS (popsaných ve třetí kapitole) v kontextu PSH a LCSH. Analýza představuje detailnější pohled na převod PSH a LCSH do datového modelu SKOS, který je společný oběma systémům, stejně jako uvádí příklady interoperability těchto systémů.

5.1.PSH a vybrané trendy ve vývoji KOS

5.1.1. PSH v datovém modelu SKOS

PSH byl oficiálně publikován v datovém modelu SKOS v roce 2009. Hlavní formát, v němž je PSH zachycen, je MARC pro autoritní záznamy. Ten se používá v automatizovaných knihovnických systémech, ale pro webovou distribuci se nehodí. „Vzhledem k požadavku na interoperabilitu s jinými KOS měl být zvolený formát založený na standardu XML (eXtensible Markup Language). Po zvážení použití formátu MADS (Metadata Authority Description Schema) pro autoritní záznamy a možnosti napsat si vlastní XML schéma byl navržen formát SKOS“ [MYNARZ; KAMRÁDKOVÁ; KOŽUCHOVÁ, 2009]. Inspirací pro převod a způsob distribuce PSH ve formátu SKOS byly již existující aplikace SKOS v jiných knihovnách. Jako jedna z prvních knihoven použila SKOS pro své věcné autority Švédská národní knihovna⁵⁶. Jiným příkladem byl Tezaurus pro ekonomii Německé národní ekonomické knihovny⁵⁷. Hlavním rozhodujícím faktorem však bylo zveřejnění hesláře v souladu se čtyřmi principy linked data, což v případě PSH zahrnuje:

- Použití URI (Uniform Resource Identifier) jako identifikátorů hesel.
- Použití HTTP (Hypertext Transfer Protocol) URI, takže je možné si hesla prohlédnout pomocí webového prohlížeče.
- Poskytnutí užitečných informací u hesla.
- Zahrnutí odkazů na další zdroje [MYNARZ; KAMRÁDKOVÁ; KOŽUCHOVÁ, 2009].

⁵⁶ <http://www.kb.se/>

⁵⁷ <http://www.zbw.eu/index-e.html>

Převod PSH do datového modelu SKOS

Při převodu PSH do formátu SKOS byly použity záznamy PSH v OAI-XML, které jsou zpřístupňovány přes rozhraní Aleph X Server. Sémanticky se tedy jednalo o formát MARC, formálně o správně utvořené XML. Díky tomu bylo možné použít jazyk XPath, který slouží pro hledání v XML dokumentech. Prvním krokem při převodu bylo určení reprezentace polí MARC záznamů ve formátu SKOS. V praxi to znamenalo na příklad to, že příbuzné heslo, které se v autoritním záznamu hesla PSH ve formátu MARC nachází v podpoli \$a pole 550 bylo převedeno do formátu SKOS jako <skos:related>. Tímto způsobem byla namapována i další pole. Následně byl vytvořen skript, který na základě systémového čísla získal z Aleph X Serveru záznam v OAI-XML, z něhož podle stanovených XPath výrazů vyextrahoval potřebné údaje. Po dalším zpracování byly tyto údaje vloženy do šablony SKOS záznamu. Nakonec bylo k souboru záznamů vytvořeno schéma, zaznamenávající strukturu PSH pomocí jeho základních 44 hesel, a také metadata ve formátu Dublin Core [MYNARZ, 2009].

PSH/MARC používá k linkování stanovená záhlaví, zatímco koncepty SKOS (<skos:concept>) jsou vzájemně propojeny použitím jednotných identifikátorů URI. Každý autoritní záznam hesla PSH vytvořený NTK zahrnuje kontrolní číslo ve formátu MARC 21 v poli 001. Tato kontrolní čísla jsou vytvářena tak, aby byla trvalá a jedinečná. Každá kontrolní číslice se skládá z alfanumerického kódu, který je tvořen tříznakovou zkratkou "PSH", za níž následuje unikátní číselná kombinace, na příklad "PSH5450". Toto číslo se stalo kandidátem pro identifikaci konceptů SKOS, tedy základem pro jednotlivá URI hesel (viz příklad č. 1, zvýrazněná část).

příklad č. 1:

```
<skos:Concept rdf:about="http://psh.ntkcz.cz/skos/PSH5450">
```

Jelikož byl SKOS vytvořen pro multi-jazykové prostředí, uživatelé SKOS jsou podporováni, aby užívali atributy pro určení jazyka. PSH je dvojjazyčný, a proto má každé heslo českou i anglickou podobu (viz příklad č. 2).

příklad č. 2:

```
<skos:prefLabel xml:lang="cs">chemie</skos:prefLabel>  
<skos:prefLabel xml:lang="en">chemistry</skos:prefLabel>
```

Formát MARC 21 pro autority rozlišuje preferovaná (1XX pole) a nepreferovaná (4XX pole) znění hesel. Podobně také SKOS slovník poskytuje dvě charakteristiky: `<skos:prefLabel>` a `<skos:altLabel>`, což umožňuje jejich přímé namapování (viz příklad č. 3).

příklad č. 3:

```
<skos:prefLabel xml:lang="cs">relativita</skos:prefLabel>  
<skos:altLabel xml:lang="cs">teorie relativity</skos:altLabel>
```

Pro vazby mezi jednotlivými preferovanými zněními se v PSH/MARC používají pole 5XX. Sémantické vztahy (vztahy hierarchie a asociace) vyjádřené ve formátu MARC jsou snadno převoditelné do formátu SKOS, který je definuje pomocí `<skos:related>`, `<skos:broader>`, `<skos:narrower>` (viz příklad č. 4).

příklad č. 4:

```
<skos:related rdf:resource="http://psh.ntkcz.cz/skos/PSH423"/>  
<skos:broader rdf:resource="http://psh.ntkcz.cz/skos/PSH3423"/>  
<skos:narrower rdf:resource="http://psh.ntkcz.cz/skos/PSH3474"/>
```

Převedení a distribuce PSH ve formátu SKOS má mnoho výhod. SKOS je jako doporučení W3C podporovaný řadou softwarových nástrojů z funkčně odlišných oblastí. Jedním z užitečných nástrojů jsou tzv. validátory, které zajišťují automatickou revizi konsistence hesláře. Validátory kontrolují vztahy vyplývající z logiky daného formátu, tedy správnost strukturního provázání hesel, na příklad nemá-li jedno heslo více nadřazených hesel apod⁵⁸. Kontrola pomocí validátoru je velmi rychlá a nenáročná. Jiným příkladem využití PSH ve formátu SKOS jsou vizualizační nástroje převádějící SKOS do jeho grafické reprezentace, která je nejen uživatelsky atraktivní, ale velmi často napomáhá v orientaci při logickém procházení struktury hesláře. Další možností je využití PSH jako řízeného slovníku při automatické indexaci dokumentů. SKOS je akceptován jako vstupní formát při indexačním procesu a výstupem v podobě dokumentu s přiřazenými předmětovými hesly. Velkou výhodou formátu SKOS je v tomto případě jeho multijazyková podpora. Při použití dobře zvoleného indexačního nástroje je možné indexovat nejen nejběžnější dokumenty anglické, ale například i texty

⁵⁸ Validační službu pro KOS ve formátu SKOS poskytuje na příklad software PoolParty (<http://poolparty.punkt.at/>).

české. A to nejen na základě statistiky slov a slovních spojení, ale i učícími algoritmy s pokročilou analýzou přirozeného jazyka [KOŽUCHOVÁ; ŠKUTA, 2010]. Proces automatické indexace využívající PSH je v současné době v Národní technické knihovně ve fázi testování.

5.1.2. Interoperabilita PSH

V kapitole 3.1 byla jako jeden z trendů ve vývoji KOS identifikována interoperabilita. SKOS je aplikací RDF, který dobře definuje logické vlastnosti. Díky tomu jsou kontrolované strukturované slovníky používající SKOS strojově srozumitelné, např. počítačová aplikace umí číst, dávat význam a poskytovat jim různé funkce. RDF podporuje distribuované publikování dat. To znamená, že strukturované kontrolované slovníky zveřejněné ve formátu SKOS mohou být odkazovány na další datové zdroje (např. další KOS). Jedná se tedy o možnost vyjádřit shodu nebo podobnost mezi pojmy různých KOS. Tyto vazby mohou být vyjádřeny hierarchicky <skos:broadMatch>, <skos:narrowMatch> nebo asociativně <skos:relatedMatch>, <skos:closeMatch>, <skos:exactMatch>. Na příklad pomocí prvku <skos:exactMatch> lze odkázat na ekvivalentní pojem v jiném pojmovém KOS, na který lze jednoznačně odkazovat s použitím URI. Tuto vlastnost demonstrují vazby hesel PSH na Předmětová hesla Kongresové knihovny a DBPedia⁵⁹ (viz příklad č. 5).

příklad č. 5:

```
<skos:exactMatch rdf:resource="http://id.loc.gov/authorities/sh85112496#concept"/>  
  <skos:exactMatch rdf:resource="http://dbpedia.org/resource/Anthropology"/>  
  <skos:closeMatch rdf:resource="http://dbpedia.org/resource/Ethnography"/>
```

Vzhledem k tomu, že lze PSH v datovém modelu SKOS nyní poměrně dobře integrovat s novými technologiemi, je možné jej implementovat nejen v knihovnách, paměťových a informačních institucích a dalších relevantních subjektech, ale i v dalších moderních

⁵⁹ DBpedia je projekt, jehož cílem je extrakce strukturovaných informací z Wikipedie za účelem jejich zpřístupnění na webu. Znalostní báze DBpedia je multijazyčná a v současnosti se skládá z více než miliardy RDF tripletů, které byly extrahovány z 92 jazykových verzí Wikipedie např. z anglické, německé, francouzské, španělské, italské, ale i české verze Wikipedie. Popisuje přes 3,4 milionů entit zahrnující osoby, místa, společnosti, nemoci apod. Pro každou z těchto entit DBpedia definuje globálně unikátní identifikátor [KOŽUCHOVÁ; ŠKUTA, 2010].

systémech. Příkladem takového systému je portál cs.wikipedia.org, v jehož rámci by bylo možné vzájemné propojení hesel PSH s hesly a nebo kategoriemi v tomto portálu⁶⁰.

V roce 2011 navázal referát PSH za tímto účelem kontakt s pracovníky české Wikipedie. Již v počátcích jednání se však narazilo na licenční problémy. Autorská díla, kterým je také PSH, jsou na portálu cs.wikipedia.org zveřejňována pod podmínkami licence Creative Commons, konkrétně varianty Uveďte autora-Zachovejte licenci 3.0 Česko (Attribution-ShareAlike 3.0 Unported, CC BY-SA 3.0 Unported, dále jen CC BY-SA). Podstatným rysem této veřejné licence je, že dílo zveřejněné na cs.wikipedia.org je možno užívat i komerčně. Tato veřejná licence je ze své podstaty bezúplatná. V této modifikaci nicméně nevylučuje užití, jehož primárním účelem je získání obchodního prospěchu⁶¹. Modifikované dílo tedy musí být poskytováno pod licencí CC BY-SA nebo obdobnou, zároveň ale není zakázáno jeho komerční užití (např. prodej nosičů s modifikovaným dílem obsahujícím PSH). NTK byla názoru, že v současné době je nemožné využít PSH k přímému hospodářskému prospěchu, a že jeho uvolnění ve prospěch široké veřejnosti nelze pojímat jako újmu. Z toho důvodu bylo PSH pod touto licencí uvolněno⁶². Na vzájemném propojení PSH a Wikipedie se tak v současné době v referátu PSH pracuje.

Další propojení PSH s jinými systémy či aplikacemi bude záležet nejen na Národní technické knihovně, jako na tvůrci PSH, ale především se bude odvíjet od poptávky

⁶⁰ Podobné oboustranné propojení již v českém prostředí existuje. Jedná se o propojení jmenných autoritních záznamů Národní knihovny ČR s hesly (články) ve Wikipedii o těchto autorech, na příklad autoritní záznam o Karlu Čapkovi http://aleph.nkp.cz/F/?func=direct&doc_number=000004031&local_base=AUT) je v katalogu po kliknutí na "Další informace" nebo na ikonické zobrazení (logo Wikipedie) prolinkován do portálu cs.wikipedia.org na článek o Karlu Čapkovi (http://cs.wikipedia.org/wiki/Karel_%C4%8Capek). Z portálu cs.wikipedia.org zase naopak funguje propojení na seznam děl daného autora v databázi Národní knihovny ČR.

⁶¹ Faktická komerční využitelnost veřejných licencí šířených za podmínky SA (Share Alike) je však omezena – jakékoliv dílo, které by obsahovalo PSH (tedy jej modifikovalo), musí být dále šířeno pod stejnou CC licencí (CC BY-SA) anebo licencí obdobnou.

⁶² Národní technická knihovna umožňuje PSH/SKOS zdarma a libovolně používat, kopírovat, šířit a upravovat, avšak s povinností uvést primárního autora (Národní technická knihovna v Praze) a při vytvoření odvozeniny z původního díla (PSH ve formátu SKOS) je nutné šířit jej pod stejnou nebo slučitelnou licencí.

uživatelů a dalšího vývoje a trendů v oblasti KOS, které jsou za tímto účelem sledovány.

5.2.LCSH a vybrané trendy ve vývoji KOS

5.2.1. LCSH v datovém modelu SKOS

Již od roku 1902 je posláním oddělení katalogizačních distribučních služeb v Kongresové knihovně Spojených států amerických umožnit knihovnám v Americe, ale i v celém světě užití a obohacování bibliografických dat. Katalogizace knihovních materiálů typicky zahrnuje dvě široké oblasti: jmennou a předmětovou katalogizaci. Zatímco jmenná katalogizace zajišťuje údržbu jednotek v katalogu, předmětová katalogizace zajišťuje údržbu různých typů KOS. V Kongresové knihovně se tato správa týká LCSH a Třídění Kongresové knihovny (Library of Congress Classification, LCC). „*Předmětová hesla Kongresové knihovny patří k nejstarším a nejznámějším systémům organizace znalostí. Jedná se o propracovaný, vyspělý systém, jehož vyjádření ve strojem srozumitelné podobě, ve webovém standardu jako je SKOS, zlepší výsledky vyhledávačů, odhalí vztahy hierarchie mezi termíny a zvýší efektivnost nástrojů ke generování metadat*“ [HARPER, 2006].

Převod LCSH do datového modelu SKOS

Kongresová knihovna typicky poskytuje LCSH pro počítačové zpracování ve formátu MARC21, ale i v MARC/XML. Harper [2006] uvedl jako jednu z možností, jak převést MARC/XML do formátu SKOS, užití tabulek XLST. Ty nabízí možnost převodu jednoho XML formátu do jiného. Místo tabulek XLST však byla pro převod MARC/XML do formátu SKOS použita pymarc library⁶³, která poskytuje objektově orientované rozhraní pro zpracování MARC/XML. Proces převodu měl dvě fáze. První zahrnovala tvorbu konceptů SKOS a jednotlivých URI, zatímco v druhé fázi došlo k jejich vzájemnému propojení [SUMMERS, 2008].

Každý autoritní záznam ve formátu MARC21, který vytváří Kongresová knihovna obsahuje v poli 001 kontrolní číslo Kongresové knihovny (Library of Congress Control Number, LCCN). Tyto kontrolní čísla jsou persistentní a jedinečné, a proto se staly základem pro rozlišení jednotlivých konceptů SKOS, k jejichž identifikaci se používají

⁶³ <http://pypi.python.org/pypi/pymarc>

URI. LCCN se skládá z alfanumerického kódu, který je tvořen dvouznakovou zkratkou "sh", za níž následuje unikátní číselná kombinace, na příklad "sh85014225". LCCN, která nemají podobu URI, byly vnořeny jako jedna jeho část podle schématu <http://lcs.info/{lccn}#concept> (viz příklad č. 1, zvýrazněná část).

příklad č. 1:

```
<rdf:Description rdf:about="http://id.loc.gov/authorities/sh85014225#concept">
```

Formát MARC pro autority rozlišuje preferovaná (1XX pole) a nepreferovaná (4XX pole) znění hesel. Podobně také SKOS slovník poskytuje dvě charakteristiky: `<skos:prefLabel>` a `<skos:altLabel>`, což umožňuje jejich přímé namapování (viz příklad č. 2).

příklad č. 2:

```
<skos:prefLabel xml:lang="en">Anatomy</skos:prefLabel>
<skos:altLabel xml:lang="en">Generation</skos:altLabel>
```

Pro vazby mezi jednotlivými předmětovými hesly se v LCSH/MARC používají pole 5XX. Sémantické vztahy vyjádřené ve formátu MARC jsou snadno převoditelné do formátu SKOS, který je definuje pomocí `<skos:related>`, `<skos:broader>`, `<skos:narrower>` (viz příklad č. 3).

příklad č. 3:

```
<skos:related rdf:resource="http://id.loc.gov/authorities/sh85090222#concept"/>
<skos:broader rdf:resource="http://id.loc.gov/authorities/sh85076841#concept"/>
<skos:narrower rdf:resource="http://id.loc.gov/authorities/sh85014225#concept"/>
```

LCSH/MARC zahrnuje samozřejmě také celou řadu údajů, které sice jsou relevantní pro Kongresovou knihovnu, ale ve formátu SKOS se prvky k jejich reprezentaci nevyskytují. Patří mezi ně na příklad informace o tom, kdy byl záznam vytvořen a kdy byl naposledy modifikován. Ačkoli takové možnosti SKOS nenabízí, problém byl vyřešen pomocí flexibility RDF, která umožňuje jiným slovníkům jako je Dublin Core (`dcterms:created`, `dcterms:modified`) tyto údaje importovat a začlenit do popisu SKOS [SUMMERS, 2008]. Summers [2008], stejně jako Harper [2006] nicméně konstatují, že značné množství informací je v procesu konverze LCSH do formátu SKOS ztraceno.

5.2.2. Interoperabilita LCSH

Příkladem propojení dvou strukturně rozdílných systémů představuje propojení LCSH a LCC. Po mnoho let bylo snahou Kongresové knihovny Spojených států amerických poskytovat linky mezi LCSH a LCC tam, kde to bylo vhodné. Zpočátku spočívaly linky pouze v LCSH, ale nyní jsou prolinkovány také autoritní záznamy LCC. Přibližně 36% předmětových hesel obsahuje číslo třídy LCC, které obecně reprezentuje běžný aspekt předmětu. V případě, že se pro jedno heslo vyskytuje několik aspektů vyjádřitelných různými čísly tříd LCC, je přidán vysvětlující termín, který indikuje specifickou disciplínu. Číslo tříd se k předmětovým heslům přidávají pouze tehdy, existuje-li blízká korelace mezi předmětovými hesly a tabulkami LCC. LCC číslo se v tištěné verzi LCSH nachází pod platným předmětovým heslem. V rámci klasifikačního webu si uživatelé mohou zobrazovat LCSH spolu s LCC [LIBRARY OF CONGRESS, 1997].

Jak již bylo řečeno, jedním z užitečných prvků datového modelu SKOS je možnost vyjádřit shodu nebo podobnost mezi pojmy různých KOS. Tyto vazby mohou být vyjádřeny hierarchicky <skos:broadMatch>, <skos:narrowMatch> nebo asociativně <skos:relatedMatch>, <skos:closeMatch>, <skos:exactMatch>. Na příklad pomocí prvku <skos:exactMatch> lze odkázat na ekvivalentní pojem v jiném pojmovém KOS, na který lze jednoznačně odkazovat s použitím URI. V případě LCSH se jedná na příklad o vazbu na RAMEAU⁶⁴ (viz příklad č. 4).

příklad č. 4:

```
<skos:closeMatch rdf:resource="http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb119736042"/>
```

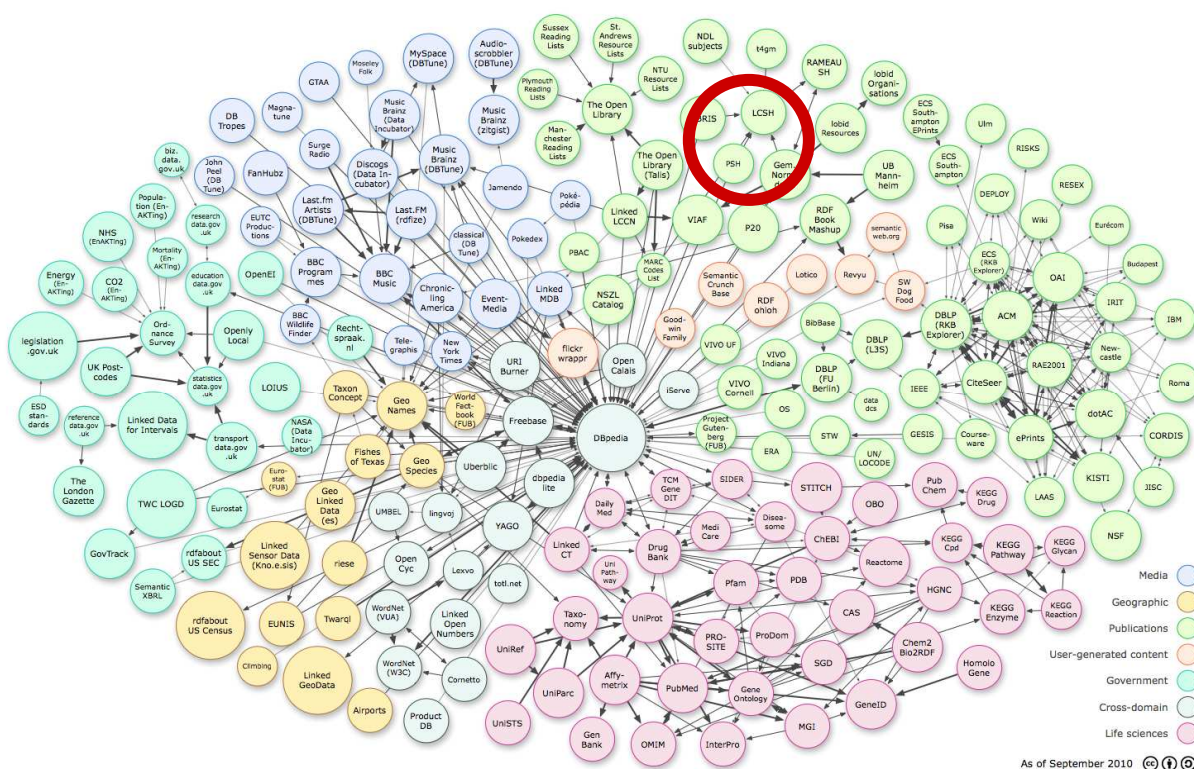
LCSH je stejně jako PSH zveřejněn v datovém modelu SKOS, a proto může využívat stejných výhod vyplývajících z aplikace RDF dobře definujícího logické vlastnosti a podporující distribuované publikování dat. To znamená, že LCSH může být nyní, podobně jako PSH, odkazován (mapován) na další datové zdroje (např. další KOS) a může také být implementován nejen v knihovnách, paměťových a informačních institucích a jiných relevantních subjektech, ale i v dalších moderních systémech. Proto

⁶⁴ RAMEAU (Répertoire d'autorité-matière encyclopédique et alphabétique unifié) je kontrolovaný slovní, podobný LCSH, který se užívá k předmětové indexaci v Národní knihovně Francie, ale i v dalších akademických, veřejných a výzkumných knihovnách, stejně jako se užívá v několika soukromých organizacích.

je na tvůrcích LCSH, jakým dalším směrem se bude celý systém vyvíjet, jakým způsobem bude využíván, popřípadě na jaké další systémy bude odkazovat.

5.3.LOD cloud diagram

„Jedna z výhod sémantického webu a linked data je, že tradičně izolované datové soubory mohou být integrovány“ [SUMMERS, 2008]. The Linking Open Data cloud diagram (dále jen LOD cloud diagram)⁶⁵ je diagram nebo také mapa, která zobrazuje takové datové soubory, které byly publikovány ve formátu linked data [CYGANIAK; JENTZSCH, 2010]. Diagram je založen na sesbíraných metadatech a každý datový soubor je prolinkován nejen se svou domovskou stránkou, ale navíc zobrazuje skutečné vazby jednotlivých datových souborů na jiné. V LOD cloud diagramu našly své místo jak LCSH, tak PSH (viz vyznačení v obr. č. 18). LOD cloud diagram udržuje Richard Cyganiak a Anja Jentzsch. Poslední aktualizace LOD cloud diagramu pochází z 22.září 2010.



Obr. 18: LOD cloud diagram [CYGANIAK; JENTZSCH, 2010]

⁶⁵ <http://richard.cyganiak.de/2007/10/lo/>

ZÁVĚR

Jak již bylo řečeno v úvodu, slovníky, tezaury, předmětová hesla, klasifikační schémata, ale i různé autoritní soubory a jiné typy systémů organizace znalostí jsou v knihovnách vytvářeny již celá desetiletí. K tradičním představitelům systémů organizace znalostí patří PSH a LCSH. Ačkoli by se mohlo zdát, že jsou PSH a LCSH neporovnatelné a v mnoha ohledech tomu tak skutečně je, na příklad jde-li o velikost (necelých 14 000 hesel PSH a zhruba 337 000 hesel LCSH), autoritu a personální sílu, která dané systémy spravuje, tradici a komunitu uživatelů (v případě PSH několik desítek knihoven, v případě LCSH stovky) či peníze, které jsou do systémů investované. Někteří dokonce tvrdí, že kromě LCSH neexistuje jiný systém, který by byl natolik propracovaný, s tak bohatou tradicí, rozšířením a zázemím. I přes tyto značné rozdíly, mají oba systémy některé společné rysy.

Za společný rys může být považována skutečnost, že se PSH v určitém směru přibližuje LCSH způsobem aktualizací. Jeden z procesů aktualizace PSH je export a následná analýza pole pro volně tvořený termín (obsah pole 653 v bibliografickém záznamu ve formátu MARC 21), které se používá při věcném popisu knih v Národní technické knihovně (NTK). Z tohoto způsobu aktualizace vyplývá fakt, že část hesel PSH vzniká na základě toho, co se nachází ve fondu NTK. Z tohoto pohledu se podobá aplikaci "literary warrant" využívaného pro LCSH, což znamená, že hesla PSH do určité míry, podobně jako LCSH v Kongresové knihovně, reflektují skutečná témata dokumentů vyskytující se ve fondu NTK a zároveň odráží terminologii používanou v současné literatuře.

Jejich další podobnost souvisí s vynálezem Internetu a rozvojem World Wide Web, který otevřel nové možnosti pro využití vyspělých systémů organizace znalostí, které tak nemusí být izolované v knihovnách, ale jejich preciznost a propracovanost může být využita také v prostředí World Wide Web. Klíčem k úspěchu je jejich aplikace v rámci vhodných standardů a sémantických webových technologií zaručujících jejich efektivitu v prostředí webového informačního chaosu.

Vedle sebe dnes stojí knihovny, komunita sémantického webu a různé metadatové iniciativy, které mají společný cíl, jímž je pojmenování konceptů a entit a jejich vzájemné propojování. Spolupráce mezi nimi je rovněž jedním z faktorů ovlivňujících

jednak rozvoj systémů organizace znalostí, jednak rozvoj sémantických webových technologií. Proto je potřeba, aby knihovny a knihovníci implementovali a zpřístupňovali systémy organizace znalostí ve formátech umožňujících vnoření webových technologií, které hrají klíčovou roli v naplnění konceptu sémantického webu.

PSH a LCSH jsou příklady tradičních systémů organizace znalostí, které se do prostředí webu adaptovaly. Jak Národní technická knihovna, tak Kongresová knihovna, přibližně ve stejné době a nezávisle na sobě aplikovaly na své systémy formát SKOS určený k reprezentaci různých systémů organizace znalostí, jejich sdílení a odkazování, podporující publikaci a užití těchto systémů v decentralizovaném, distribuovaném informačním prostředí jako je World Wide Web.

V budoucnu bude záležet především na tvůrcích PSH a LCSH, jakým dalším směrem se budou tyto systémy ubírat, jakým způsobem budou nabízeny, distribuovány a využívány. Není pochyb o tom, že se jejich rozvoj bude odvíjet především od poptávky uživatelů a bude odvozován od dalšího vývoje a trendů v oblasti systémů organizace znalostí.

SEZNAM POUŽITÉ LITERATURY

- AITCHISON, Jean; CLARKE, Stella Dextre. *The Thesaurus : A Historical Viewpoint, with a Look to the Future*. In ROE, Sandra K.; THOMAS, Alan R. *The Thesaurus : Review, Renaissance, and Revision*. New York : Haworth Information Press, 2004. s. 5-21. ISBN 0-7890-1978-7.
- BALÍKOVÁ, Marie. Katalogizační politika : Aplikace Library of Congress Subject Headings System v Národní knihovně ČR. *Národní knihovna : knihovnická revue* [online]. 2001, č. 4, [cit. 2011-06-28]. Dostupný z WWW: <<http://knihovna.nkp.cz/NKKR0104/0104303.html>>. ISSN 1214-0678.
- BALÍKOVÁ, Marie. Pořádání informací. In: *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online databáze]. Praha: Národní knihovna ČR, 2003- [cit. 2011-08-01]. Dostupné z: <http://aleph.nkp.cz/F/?func=direct&doc_number=000001604&local_base=KTD>.
- BERNERS-LEE, Tim . *Decentralized Information Group (DIG)* [online]. 2008 [cit. 2011-03-03]. Timbl's blog. Dostupné z WWW: <<http://dig.csail.mit.edu/breadcrumbs/blog/4>>.
- BROUGHTON, Vanda. *Essential thesaurus construction*. 1st ed. London : Facet Publishing, 2006. 296 s. ISBN 978-1-85604-565-0.
- CELBOVÁ, Ludmila. 2003. Portál. In: *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online databáze]. Praha: Národní knihovna ČR, 2003- [cit. 2011-08-03]. Dostupné z: <http://aleph.nkp.cz/F/?func=direct&doc_number=000000550&local_base=KTD>.
- CELBOVÁ, Ludmila. 2003a. Sémantický web. In: *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online databáze]. Praha: Národní knihovna ČR, 2003- [cit. 2011-07-18]. Dostupné z: <http://aleph.nkp.cz/F/?func=direct&doc_number=000000556&local_base=KTD>.
- CLARKE, Stella Dextre. ISO 25964 : a standard in support of KOS interoperability. In *ISKO UK Conference* [online]. London : ISKO UK, 2011 [cit. 2011-07-25]. Dostupné z WWW: <<http://www.iskouk.org/conf2011/papers/dextreclarke.pdf>>. Preprint.
- CYGANIAK, Richard; JENTZSCH, Anja . *The Linking Open Data cloud diagram* [online]. 2010 [cit. 2011-08-07]. The Linking Open Data cloud diagram. Dostupné z WWW: <<http://richard.cyganiak.de/2007/10/lod/>>.
- DCMI. *Dublin Core Metadata Initiative (DCMI)* [online]. 2009 [cit. 2011-02-24]. DCMI/NKOS Task Group. Dostupné z WWW: <<http://dublincore.org/groups/nkos/>>.

- Groupe de travail Interop - AFUL. *Interoperability* [online]. [2009] [cit. 2011-08-03]. Definition : Interoperability. Dostupné z WWW: <<http://interoperability-definition.info/en/>>.
- HARPER, Corey A. Encoding Library of Congress Subject Headings in SKOS : Authority Control for the Semantic Web. In *International Conference on Dublin Core and Metadata Applications, 3 - 6 October 2006* [online]. Mexico : International Conference on Dublin Core and Metadata Applications, 2006 [cit. 2011-05-31]. Dostupné z WWW: <<https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/3268/dc2006.pdf?sequence=1>>.
- HJØRLAND, Birger. What is Knowledge Organization (KO)?. *Knowledge Organization*. 2008, vol. 35, no. 2/3, s. 86-101. ISSN 0943-7444.
- HODGE, Gail. *Systems of Knowledge Organization for Digital Libraries : Beyond Traditional Authority Files*. Washington, DC : The Digital Library Federation, 2000. 35 s. Dostupný také z WWW: <<http://www.clir.org/pubs/reports/pub91/pub91.pdf>>. ISBN 1-887334-76-9.
- CHOWDHURY, G. G.; CHOWDHURY, S. 2007. *Organizing information from the shelf to the web*. 1st edition. London : Facet , 2007. 230 s. ISBN 978-1-85604-578-0.
- ISKO. *ISKO : International Society for Knowledge Organization* [online]. 2010 [cit. 2011-07-25]. About ISKO. Dostupné z WWW: <<http://www.isko.org/about.html>>.
- JOHNSON, Eric H. Distributed Thesaurus Web Services. In ROE, Sandra K.; THOMAS, Alan R. *The Thesaurus : Review, Renaissance, and Revision*. New York : Haworth Information Press, 2004. s. 121-153. ISBN 0-7890-1978-7.
- KOŽUCHOVÁ, Kristýna; ŠKUTA, Ctibor. 2010. Polytematický strukturovaný heslář a jeho potenciál v oblasti třídění a zpřístupňování webových dokumentů. In *INFORUM 2010: 16. ročník konference o profesionálních informačních zdrojích, Praha 25.-27. května 2010* [online]. Praha : Albertina icome Praha, 2010 [cit. 2010-06-01]. Text příspěvku a prezentace dostupné na World Wide Web: <<http://www.inforum.cz/sbornik/2010/67>>. ISSN 1801-2213.
- KOŽUCHOVÁ, Kristýna; ŠKUTA, Ctibor. 2010a. Vliv trendů systémů organizace znalostí na vývoj Polytematického strukturovaného hesláře v Národní technické knihovně . *Knihovna plus* [online]. 2010, č. 1 [cit. 2010-04-01]. Dostupný z World Wide Web: <<http://knihovna.nkp.cz/knihovnaplus101/skuta.htm>>. ISSN 1801-5948.
- Library of Congress. 2011. *Cataloging Distribution Service* [online]. [2011] [cit. 2011-07-21]. Library of Congress Subject Headings. Dostupné z WWW: <<http://www.loc.gov/cds/products/product.php?productID=44>>.

- Library of Congress. 2011a. *Library of Congress* [online]. 2011 [cit. 2011-06-27]. Subject Headings Weekly Editorial Meeting : Summary of Decisions. Dostupné z WWW: <<http://www.loc.gov/catdir/cpsd/summdec.html>>.
- Library of Congress. *Library of Congress Subject Headings*. 20th edition. Washington, D.C. : Library of Congress Cataloging Distribution Service, 1997. Introduction, s. vii-xiv.
- MILLER, Eric. The Semantic Web and Digital Libraries [prezentace]. In *International Conference on Dublin Core and Metadata Applications, Shanghai, China, 11-14 October 2004* [online]. Shanghai: DC2004, 2004 [cit. 2011-03-03]. Prezentace dostupná na World Wide Web: <<http://www.w3.org/2004/Talks/1013-semweb-em/talk>>.
- MILES, Alistair. The Simple Knowledge Organization System (SKOS) in the context of Semantic Web Deployment [prezentace]. In *Library of Congress, May 2008* [online]. [cit. 2011-03-03]. Prezentace dostupná na World Wide Web: <<http://www.slideshare.net/gardensofmeaning/simple-knowledge-organization-system-skos-in-the-context-of-semantic-web-deployment-library-of-congress-may-2008>>.
- MILES, Alistair; et al. SKOS : a language to describe simple knowledge structures for the Web. In *XTech 2005: XML, the Web and beyond* [online]. [cit. 2011-03-03]. Dostupný z WWW: <<http://www.webcitation.org/5m2gMRJih>>.
- Networked Knowledge Organization Systems/Services. *NKOS Networked Knowledge Organization Systems and Services* [online]. 2000 [cit. 2011-02-10]. Taxonomy of Knowledge Organization Sources/Systems. Dostupné z WWW: <http://nkos.slis.kent.edu/KOS_taxonomy.htm>.
- MYNARZ, Jindřich. Jak lze prakticky využít Polytematický strukturovaný heslář pro věcný popis elektronických zdrojů. *Ikaros* [online]. 2009, roč. 13, č. 12 [cit. 2009-12-01]. Dostupný na World Wide Web: <<http://www.ikaros.cz/node/5872>>. URN-NBN:cz-ik5591. ISSN 1212-5075.
- MYNARZ, Jindřich. *Jak se převáděl PSH do SKOS*. 2009. 2 s. Interní materiál pro potřeby Národní technické knihovny.
- MYNARZ, Jindřich; KAMRÁDKOVÁ, Kateřina; KOŽUCHOVÁ, Kristýna. Novinky z oblasti Polytematického strukturovaného hesláře. *Ikaros* [online]. 2009, roč. 13, č. 7 [cit. 2009-07-01]. Dostupný na World Wide Web: <<http://www.ikaros.cz/node/5591>>. URN-NBN:cz-ik5591. ISSN 1212-5075.
- MYNARZ, Jindřich; ZEMÁNEK, Jan. Úvod k linked data. *Knihovna plus* [online]. 2010, č. 1 [cit. 2011-08-02]. Dostupný z WWW: <<http://knihovna.nkp.cz/knihovnaplus101/myna.htm>>. ISSN 1801-5948.
- PASTOR-SANCHEZ, Juan-Antonio; MARTINEZ MENDEZ, Francisco Javier; RODRÍGUEZ-MUÑOZ, José Vicente. Advantages of thesaurus representation using the Simple Knowledge Organization System (SKOS) compared with

proposed alternatives. *Information Research : an international electronic journal* [online]. 2009, vol. 14, no. 4 [cit. 2011-03-23]. Dostupný z WWW: <<http://informationr.net/ir/14-4/paper422.html>>. ISSN 1368-1613.

- PHIPPS, Jon. SKOS: A W3C Recommendation [prezentace]. In *ASIS&T 2010, Pittsburgh, PA, USA, 2010* [online]. [cit. 2011-03-23]. Prezentace dostupná na World Wide Web: <<http://nkos.slis.kent.edu/2010ASIST/ASIST2010-2-SKOS-Phipps.pdf>>.
- QIN, Jian . Representation and Organization of Information in the Web Space : From MARC to XML. *Informing Science*. 2000, vol. 3, no. 2, s. 83-87. Dostupný také z WWW: <<http://inform.nu/Articles/Vol3/v3n2p83-88.pdf>>. ISSN 1547-9684.
- SAUMURE, Kristie; SHIRI, Ali. Knowledge organization trends in library and information studies : a preliminary comparison of the pre- and post-web areas. *Journal of information science*. 2008, vol. 35, no. 5, s. 651-666. ISSN 0165-5515.
- SHIRI, Ali; REVIE, Crawford. Thesauri on the Web : current developments and trends. *Online Information Review*. 2000, vol. 24, no. 4, s. 273-279. ISSN 1468-4527.
- SKLENÁK, Vilém, et al. *Data, informace, znalosti a Internet*. 1. vyd. Praha : C. H. Beck, 2001. 507 s. ISBN 80-7179-409-0.
- SKLENÁK, Vilém. WWW. In: *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online databáze]. Praha: Národní knihovna ČR, 2003- [cit. 2011-07-18]. Dostupné z: <http://aleph.nkp.cz/F/?func=direct&doc_number=000000687&local_base=KTD>.
- SKOLKOVÁ, Linda. Polytematický strukturovaný heslář (Polythematic Structured Subject Heading System). Praha, 2007. iv, 172 s., 27 s. příl. + 1 CD-ROM. Diplomová práce. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví. Vedoucí diplomové práce Ing. Miloslav Nič, PhD. Dostupný také na WWW: <http://www.skolkova.net/diplomka/SKOLKOVA_Linda_DP_PSH.pdf>.
- SMOLKA, Pavel. Polytematický strukturovaný heslář. *Ikaros* [online]. 1999, roč. 3, č. 8 [cit. 07.04.2011]. Dostupný na World Wide Web: <<http://www.ikaros.cz/node/1030>>. URN-NBN:cz-ik1030. ISSN 1212-5075.
- SMRT, Martin. Proč analyzovat logy. *Newsletter Dobrý tip* [online]. 2007, č. 5, [cit. 2011-07-19]. Dostupný z WWW: <<http://www.dobryweb.cz/newsletter-proc-analyzovat-logy/>>.
- SOUZA, Renato Rocha; TUDHOPE, Douglas; ALMEIDA, Maurício Barcellos. The KOS spectra : a tentative typology of Knowledge Organization Systems. In *ISKO UK* [online]. 2010 [cit. 2011-02-10]. Dostupné

z WWW: <<http://mba.eci.ufmg.br/downloads/ISKO%20Rome%202010%20submitted.pdf>>.

- SUMMERS, Ed, et al. LCSH, SKOS and Linked Data. In *DC 2008 : International Conference on Dublin Core and Metadata Applications, Berlin 22.- 26. September 2008*. Berlin : DCMI, 2008 [cit. 2010-07-07]. Dostupný z WWW: <<http://arxiv.org/ftp/arxiv/papers/0805/0805.2855.pdf>>.
- STONE, Alva. The LCSH Century : A Brief History of the Library of Congress Subject Headings, and Introduction to the Centennial Essays. *Cataloging & Classification Quarterly*. 2000, Vol. 29, no. 1-2, s. 1. Dostupný také z WWW: <<http://catalogingandclassificationquarterly.com/ccq29nr1-2ed.htm>>.
- ŠKUTA, Ctibor. Automatizace knihovnických systémů s využitím technologie Sikuli. In *Knihovny současnosti 2010 : sborník ze 18. konference, konané ve dnech 14.-16. září 2010 v Seči u Chrudimi*. [Sestavila Eva Svobodová]. Ostrava : Sdružení knihoven ČR, 2010, s. 45 - 51. ISBN 978-80-86249-59-9.
- ŠKUTA, Ctibor; KOŽUCHOVÁ, Kristýna. Vliv trendů systémů organizace znalostí na vývoj Polytematického strukturovaného hesláře v Národní technické knihovně . *Knihovna plus* [online]. 2010, č. 1 [cit. 2011-06-27]. Dostupný z WWW: <<http://knihovna.nkp.cz/knihovnaplus101/sktua.htm>>. ISSN 1801-5948.
- TKAČÍKOVÁ, Daniela. Interoperabilita. In: *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online databáze]. Praha: Národní knihovna ČR, 2003- [cit. 2011-07-26]. Dostupné z: <http://aleph.nkp.cz/F/E9E2UN19LUF2UPS93PGIC2RTLIB99V386DYPFYGLFTQR5K3E9U-24212?func=find-aut-bib&find_request=interoperabilita&find_code=TR&find_base=KTD>.
- TUDHOPE, Douglas; NIELSEN, Marianne Lykke. Introduction to Knowledge Organization Systems and Services. *New Review of Hypermedia and Multimedia*. 2006, vol. 12, no. 1, s. 3-9. Dostupný také z WWW: <<http://www.informaworld.com/smpp/section?content=a749307110&fulltext=713240928>>. ISSN 1361-4568.
- Webify. In *WordWeb Online* [online]. Brighton (UK) : WordWeb Software, c2005-2009, last update March 2011 [cit. 2011-08-02]. Dostupný z WWW: <<http://www.wordwebonline.com/en/WEBIFY>>.
- WRIGHT, Sue Ellen. Typology for KRRs : Knowledge Representation Resources. In *NKOS-CENDI* [online]. Kent, Ohio : Kent State University, 2008 [cit. 2011-02-10]. Dostupné z WWW: <<http://nkos.slis.kent.edu/2008workshop/SueEllenWright.pdf>>.
- Státní technická knihovna. *Zásady pro vytváření Polytematického strukturovaného hesláře (PSH) : upravená verze*. Praha : Státní technická knihovna, prosinec 1996. 13 s.

- ZENG, Marcia Lei. Knowledge Organization Systems (KOS). *Knowledge Organization* . 2008, vol. 35, no. 2/3, s. 160-186. ISSN 0943-7444.
- ZENG, Marcia Lei; CHAN, Lois Mai. Trends and issues in establishing interoperability among knowledge organization systems. *Journal of the American Society for Information Science and Technology (JASIST)*. 2004, vol. 55, i. 5, s. 377-395. Dostupný také komerčně z WWW: <<http://onlinelibrary.wiley.com/doi/10.1002/asi.10387/full>>. ISSN 1532-2882.

SEZNAM POUŽITÝCH ZKRATEK

DNS	Domain Name Service
HTTP	Hypertext Transfer Protocol
LCC	Třídění Kongresové knihovny
LCCN	Library of Congress Control Number
LCSH	Předmětová hesla Kongresové knihovny
LISTA	Library, Information Science, and Technology Abstracts
MACS	Multilingual Access to Subjects
MARC	Machine-Readable Cataloging
MK	Ministerstvo kultury ČR
NTK	Národní technická knihovna
OWL	Web Ontology Language
PSH	Polytematický strukturovaný heslář
RDF	Resource Description Framework
SKOS	Simple Knowledge Organization System
STK	Státní technická knihovna
URI	Uniform Resource Identifier
XML	Extensible Markup Language

SEZNAM OBRÁZKŮ

OBR. 1: TAXONOMIE KOS PODLE SOUZY; TUDHOPEA; ALMEIDY [2010].....	15
OBR. 2: TAXONOMIE KOS A ZDROJE REPREZENTACE ZNALOSTÍ PODLE WRIGHT [2008]	16
OBR. 3: TAXONOMIE KOS PODLE ZENG [2008].....	18
OBR. 4: NÁVRH PARAMETRŮ HODNOCENÍ PODLE AUTORŮ SOUZY, TUDHOPEA A ALMEIDY [2010]	21
OBR. 5: ZOBRAZENÍ KATEGORIÍ (ZÁLOŽKA "ZÁBAVA") NABÍZENÝCH PROSTŘEDNICTVÍM WEBOVÉHO PORTÁLU YAHOO!	30
OBR. 6: GRAFICKÉ ZNÁZORNĚNÍ INTEROPERABILITY [GROUPE, 2009]	37
OBR. 7: PŘÍKLAD ZOBRAZENÍ ZÁZNAMU HESLA "JADERNÁ FYZIKA" VE FORMÁTU MARC 21 PRO AUTORITY.....	46
OBR. 8: PŘÍKLAD ZOBRAZENÍ ZÁZNAMU HESLA "JADERNÁ FYZIKA" VE FORMÁTU SKOS	47
OBR. 9: PŘÍKLAD GRAFICKÉHO ZOBRAZENÍ HESLA "JADERNÁ FYZIKA" NA STRÁNKÁCH "PROHLÍŽENÍ HESLÁŘE"	48
OBR. 10: PŘÍKLAD ZOBRAZENÍ HESLA "ASTRONOMICKÉ PŘÍSTROJE" V ROZHRANÍ PSH MANAGER ONLINE	49
OBR. 11: WEBOVÝ FORMULÁŘ PRO ZASÍLÁNÍ NÁVRHŮ NOVÉHO HESLA PSH	53
OBR. 12: PŘÍKLAD ZOBRAZENÍ ZÁZNAMU HESLA "NUKLEÁRNÍ ENERGIE" (NUCLEAR ENERGY) VE FORMÁTU MARC 21 PRO AUTORITY V ROZHRANÍ KLASIFIKAČNÍHO WEBU.....	56
OBR. 13: PŘÍKLAD ZOBRAZENÍ ZÁZNAMU HESLA "JADERNÝ PRŮMYSL" (NUCLEAR INDUSTRY) VE FORMÁTU SKOS	57
OBR. 14: KLASIFIKAČNÍ WEB: HLAVNÍ MENU	58
OBR. 15: PŘÍKLAD ZOBRAZENÍ ZÁZNAMU HESLA "JADERNÁ ENERGIE" (NUCLEAR ENERGY) V ROZHRANÍ KLASIFIKAČNÍHO WEBU	59
OBR. 16: PŘÍKLAD VIZUALIZACE HESLA "JADERNÁ ENERGIE" (NUCLEAR ENERGY).....	60
OBR. 17: WEBOVÝ FORMULÁŘ PRO ZASÍLÁNÍ NÁVRHŮ NA ZMĚNY NEBO NOVÁ PŘEDMĚTOVÁ HESLA LCSH	62
OBR. 18: LOD CLOUD DIAGRAM [CYGANIAK; JENTZSCH, 2010]	72

EVIDENCE VÝPŮJČEK

Prohlášení:

Dávám svolení k půjčování této diplomové práce. Uživatel potvrzuje svým podpisem, že bude tuto práci řádně citovat v seznamu použité literatury.

V Praze, 18. srpna 2011

Kristýna Kožuchová

[illegible]